

11-2014

Vol. 13, No. 2 (Full Issue)

JMASM Editors

Follow this and additional works at: <http://digitalcommons.wayne.edu/jmasm>



Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Editors, JMASM (2014) "Vol. 13, No. 2 (Full Issue)," *Journal of Modern Applied Statistical Methods*: Vol. 13 : Iss. 2 , Article 33.

DOI: 10.22237/jmasm/1414816320

Available at: <http://digitalcommons.wayne.edu/jmasm/vol13/iss2/33>

This Full Issue is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized editor of DigitalCommons@WayneState.

Vol. 13, No. 2 (Full Issue)

Erratum

In first article of the original published version of this issue, Panels B and C of Figure 15, p. 38, were reversed. This error has been corrected.

```
do i1 = 1,4
  j(1) = i1
  do i2 = 1,4
    j(2) = i2
    do i3 = 1,4
      j(3) = i3
      do i4 = 1,4
        j(4) = i4
        if (j(1) .eq. j(2) .or. j(1) .eq. j(3) .or. j(1) .eq. j(4)) cycle
        if (j(2) .eq. j(3) .or. j(2) .eq. j(4)) cycle
        if (j(3) .eq. j(4)) cycle
        print*,j(1),j(2),j(3),j(4)
      end do
    end do
  end do
end do
```

Journal of Modern Applied Statistical Methods

Invited Articles

Joel R. Levin, John M. Ferron,
and Boris S. Gafurov

Vol. 13, No. 2 • November, 2014

Journal of Modern Applied Statistical Methods

Shlomo S. Sawilowsky

SENIOR EDITOR

College of Education
Wayne State University

Jack Sawilowsky

EDITOR

Reason Statistical Consulting

Harvey Keselman

ASSOCIATE EDITOR EMERITUS

Department of Psychology
University of Manitoba

Alan Klockars

ASSISTANT EDITOR EMERITUS

Educational Psychology
University of Washington

Bruno D. Zumbo

ASSOCIATE EDITOR

Measurement, Evaluation,
& Research Methodology
University of British Columbia

Vance W. Berger

ASSISTANT EDITOR

Biometry Research Group
National Cancer Institute

Todd C. Headrick

ASSISTANT EDITOR

Educational Psychology
& Special Education
So. Illinois University–
Carbondale

Julie M. Smith, PhD

EDITORIAL ASSISTANT

Joshua Neds-Fox

EDITORIAL ASSISTANCE

JMASM (ISSN 1538–9472, <http://digitalcommons.wayne.edu/jmasm>) is an independent, open access electronic journal, published biannually in May and November by JMASM Inc. (PO Box 48023, Oak Park, MI, 48237) in collaboration with the Wayne State University Library System. *JMASM* seeks to publish (1) new statistical tests or procedures, or the comparison of existing statistical tests or procedures, using computer-intensive Monte Carlo, bootstrap, jackknife, or resampling methods, (2) the study of nonparametric, robust, permutation, exact, and approximate randomization methods, and (3) applications of computer programming, preferably in Fortran (all other programming environments are welcome), related to statistical algorithms, pseudo-random number generators, simulation techniques, and self-contained executable code to carry out new or interesting statistical methods.

Journal correspondence (other than manuscript submissions) and requests for advertising may be forwarded to ea@jmasm.com. See back matter for instructions for authors.

Journal of Modern Applied Statistical Methods

Vol. 13, No. 2

❧ November 2014 ❧

Table of Contents

Invited Articles

2 – 52	J. R. LEVIN J. M. FERRON B. S. GAFUROV	Improved Randomization Tests for a Class of Single Case Intervention Designs
--------	---	---

Regular Articles

53 – 75	J. R. CHEEMA	Some General Guidelines for Choosing Missing Data Handling Methods In Educational Research
76 – 83	D. J. BEST J. C. W. RAYNER	Conover's F Test as an Alternative to Durbin's Test
84 – 105	H. FINCH	A Comparison of Methods for Group Prediction with High Dimensional Data
106 – 130	A. BEAUDUCEL F. SPOHN	Retained Components Factor Transformation: Factor Loadings and Factor Score Predictors In the Column Space of Retained Components
131 – 150	N. H. JADHAV D. N. KASHID	Robust Winsorized Shrinkage Estimators for Linear Regression Model
151 – 168	I. C. A. OYEKA C. C. NNANATU	Pairwise Comparison in Repeated Measures
169 – 184	S. SEN R. LAMICHHANE N. DIAWARA	A Bivariate Distribution with Conditional Gamma and its Multivariate Form

185 – 200	D. M. SAKATE D. N. KASHID	Comparison of Estimators in GLM with Binary Data
201 – 225	R. AL-AQTASH C. LEE F. FAMOYE	Gumbel-Weibull Distribution: Properties and Applications
226 – 243	P. L. RAMOS F. A. MOALA J. A. ACHCAR	Objective Priors for Estimation of Extended Exponential Geometric Distribution
244 – 258	B. A. PARA T. R. JAN.	Discrete Generalized Burr-Type XII Distribution
259 – 286	T. N. SINDHU N. FEROZE M. ASLAM	Bayesian Estimation of the Parameters of Two-Component Mixture of Rayleigh Distribution under Doubly Censoring
287 – 306	J. A. ADEWARA U. A. MBATA	Survival Estimation Using Bootstrap, Jackknife and K-Repeated Jackknife Methods
307 – 323	A. HASSAN M. AHMAD	Life Testing Analysis of Failure Censored Generalized Exponentiated Data
324 – 354	R. R. L. KANTAM B. SRIRAM	Some Methods of Estimation from Censored Samples in Exponential and Gamma Models
355 – 363	G. KHALAF M. IGUERNANE	Ridge Regression and Ill-Conditioning
364 – 378	D. RAHARDJA	Comparison of Individual and Moving Range Chart Combinations to Individual Charts In Terms of ARL after Designing for a Common “All OK” ARL
379 – 388	A. H. KHAN T. R. JAN	Reliability Estimates of Generalized Poisson Distribution and Generalized Geometric Series Distribution
389 – 398	A. H. KHAN T. R. JAN	Estimation of Multi Component Systems Reliability in Stress-Strength Models

399 – 419	J. ARASAN M. B. ADAM	Double Bootstrap Confidence Interval Estimates with Censored and Truncated Data
420 – 431	F. SADIA S. S. HOSSAIN	Contrast of Bayesian and Classical Sample Size Determination
432 – 443	O. M. YOUSEF S. A. AL-SUBH	Estimation of Gumbel Parameters under Ranked Set Sampling
444 – 454	M. GHAHRAMANI	The Information Criterion
455 – 463	M. RASEKHI B. JAMSHIDI F. RIVAZ	Optimal Location Design for Prediction of Spatial Correlated Environmental Functional Data
464 – 478	D. L. POSTON E. CONDE	Missing Data and the Statistical Modeling of Adolescent Pregnancy
479 – 492	O. S. YAYA O. E. OLUBUSOYE O. O. OJO	Estimates and Forecasts of GARCH Model under Misspecified Probability Distributions: A Monte Carlo Simulation Approach

Emerging Scholars

493 – 505	J. G. D'CUNHA K. A. RAO	Bayesian Inference for Volatility of Stock Prices
506 – 527	E. EDIONWE J. I. MBEGBU	Local Bandwidths for Improving Performance Statistics of Model-Robust Regression 2

Statistical Software Applications and Review

528 – 545	X. LIU	Fitting Stereotype Logistic Regression Models for Ordinal Response Variables In Educational Research (Stata)
-----------	--------	--

Invited Article: **Improved Randomization Tests for a Class of Single-Case Intervention Designs**

Joel R. Levin
University of Arizona
Tucson, AZ

John M. Ferron
University of South Florida
Tampa, FL

Boris S. Gafurov
George Mason University
Fairfax, VA

Forty years ago, Eugene Edgington developed a single-case AB intervention design-and-analysis procedure based on a random determination of the point at which the B phase would start. In the present simulation studies encompassing a variety of AB-type contexts, it is demonstrated that by also randomizing the order in which the A and B phases are administered, a researcher can markedly increase the procedure's statistical power.

Keywords: Single-case intervention research, design and statistical analysis, randomization tests, statistical power, internal validity, scientific credibility

Introduction

Single-case designs that focus on behavioral and academic interventions are prevalent in a variety of clinical and educational fields (see, for example, [Kratochwill & Levin, 2014](#)). In contrast to conventional group intervention designs, single-case designs typically include only one or a few units (e.g., individuals, small groups, classrooms) to whom the intervention is administered. In addition, single-case intervention designs are intensive and implemented over longer periods of time, with more numerous assessments of the outcome measures ([Horner & Odom, 2014](#); [Kratochwill et al., 2010](#)). Single-case intervention designs that currently incorporate formal criteria to enhance their scientific

Joel R. Levin is a Professor Emeritus in the Department of Educational Psychology. Email at jrlevin@u.arizona.edu. John M. Ferron is a Professor in the Department of Educational and Psychological Studies in the College of Education. Email at ferron@usf.edu. Boris S. Gafurov is an Assistant Professor in the College of Education and Human Development. Email at bgafurov@gmu.edu.

Erratum: In the original published version of this article, Panels B and C of Figure 15 were reversed. This error is now corrected.

credibility (Levin, 1994) include ABAB designs, alternating treatment designs, and multiple-baseline designs (Kratochwill et al., 2013).

As the methodological rigor of single-case intervention designs has evolved over the years (Kratochwill & Levin, 2010), so too have the formal statistical-analysis procedures that accompany them (see, for example, Kratochwill & Levin, 2014; and Manolov, Evans, Gast, & Perdices, 2014). Although various visual/graphical approaches remain an analytic staple of single-case data (e.g., Auerbach & Zeitlin, 2014; Kratochwill, Levin, Horner, & Swoboda, 2014; Parker, Vannest, & Davis, 2014), improved statistical methods have increasingly been considered as viable supplements to visual analysis. These improved statistical methods include econometric time-series analyses (e.g., McCleary & Welsh, 1992), adapted regression- and hierarchical linear modeling procedures (e.g., Maggin et al., 2011; Manolov & Solanas, 2013; Moeyaert, Ferron, Beretvas, Van den Noortgate, & Beretvas, 2014; Shadish, Kyse, & Rindskopf, 2013), and nonparametric permutation and randomization tests (e.g., Edgington & Onghena, 2007; Ferron & Levin, 2014; Heyvaert & Onghena, 2014). The last of these statistical approaches is the focus of the present study.

Overview of the Present Study

The motivation for single-case researchers to adopt a randomization test as one component of their analytic armament is that randomization tests provide strict control of the Type I error rate (i.e., the probability of concluding that phase-to-phase differences in level, trend, variability, etc. are present when those differences are simply chance fluctuations) as long as: (1) the design includes randomization; (2) the accompanying statistical test is conducted in a manner that is consistent with the design frame; and (3) the test statistic is chosen without knowledge of the results (Edgington, 1980; Ferron & Levin, 2014). In contrast, demonstration of Type I error control has been elusive in studies of visual analysis (e.g., Ferron & Jones, 2006; Fisch, 2001; Stocks & Williams, 1995). Moreover, with regression and hierarchical models, Type I error control hinges on a relatively strong set of assumptions (Ferron, Moeyaert, Van den Noortgate, & Beretvas, 2014). The modeling assumptions include: (1) the error distribution is correctly specified (e.g., normally distributed, homogeneous variances across phases, and a first-order autoregressive function); (2) the baseline trajectory is correctly specified; (3) the baseline trajectory can be extrapolated (i.e., had the intervention not been implemented, the baseline trajectory would have continued, implying that there were no confounding effects of external events on the time

IMPROVED RANDOMIZATION TESTS

series); and (4) the treatment phase trajectory is correctly specified. Accordingly, a single-case researcher may plan a multicomponent analysis in which visual analysis serves as the primary analysis tool, a randomization test is employed to ensure that the Type I error rate is controlled, and a regression-based or hierarchical linear model is examined to summarize and estimate the size of the effect(s).

A concern with the addition of randomization tests to the analytic plan is that such tests require the researcher to introduce randomization into the design, and if the randomization is not carefully planned it can lead to a design that falls short of single-case design standards (e.g., Ferron & Levin, 2014; Kazdin, 1980; Kratochwill et al., 2010). As a consequence, researchers are encouraged to reflect carefully on the practical constraints of the context in which the study is conducted, on the desired design features (e.g., minimum phase lengths), and then tailor the randomization strategy to meet these constraints. Restricted randomization schemes have been developed to ensure that: (1) the desired number of phases and minimum phase lengths are included in reversal designs (Onghena, 1992); (2) the treatment alternates quickly enough in an alternating treatment design (Onghena & Edgington, 1994); (3) the baseline series stabilizes prior to commencement of the intervention phase (Ferron & Ware, 1994); (4) the intervention start points are staggered by a minimum amount of time in multiple-baseline designs (Koehler & Levin, 1998), and (5) researchers are able to obtain visually acceptable patterns by extending phases in multiple-baseline designs (Ferron & Jones, 2006) and reversal designs (Ferron & Levin, 2014).

The present Monte Carlo simulation study employs nonparametric randomization tests in the company of a recently proposed methodological addition that greatly enhances the internal validity of AB and ABAB single-case intervention designs (Ferron & Levin, 2014; Levin, Evmenova, & Gafurov, 2014). In these designs, A typically represents a baseline, control, or standard treatment phase containing repeated outcome measurements and B represents an intervention, experimental, or new treatment phase also containing repeated outcome measurements. Here we examine the methodological addition's effect on the statistical conclusion validity (manifested by both Type I error control and increased statistical power) of randomization tests in single-case AB and ABAB designs, in both their single-case ($N = 1$) and multiple-case ($N > 1$) forms. In the following section, we first describe the methodological addition that enhances the internal validity (scientific credibility) of single-case intervention research and then outline how the addition is incorporated into a randomization test to improve the test's statistical conclusion validity. Our decision to start our investigations

with a single-participant ($N = 1$) AB design was not because we are advocating for the use of such a design, but because it provides the simplest point to begin study of the impact of the methodological addition. Once we have established the effects on statistical conclusion validity in the simplest situation, we will progressively add complexities to strengthen the design, building to the multiple-participant ($N > 1$) ABAB design.

Edgington's (1975) Random Intervention Start-Point Model

Of four different types of randomization that can be incorporated into randomization in single-case AB experimental studies (specifically, within-case phase randomization, between-case intervention randomization, case randomization, and intervention start-point randomization (see [Ferron & Levin, 2014](#)), the last, highly creative, type was originally developed by Edgington (1975) and requires that the researcher: (1) randomly select an intervention start point from two or more that had been previously deemed acceptable; and then (2) assign to the case the start point that was actually selected. Although not applied in the conventional treatment randomization manner, this unique form of randomization increases a single-case study's internal validity and, when accompanied by the statistical test described in the following paragraph, it can increase the study's statistical conclusion validity as well. Moreover, this randomized intervention start-point approach can function to provide a true (i.e., scientifically credible) experimental comparison of two or more intervention (or intervention and control) conditions based on either one case or multiple cases per condition (for examples and discussion, see [Ferron & Levin, 2014](#); [Koehler & Levin, 1998](#); [Levin, Lall, & Kratochwill, 2011](#); [Levin & Wampold, 1999](#); and [Marascuilo & Busk, 1988](#)).

With the randomized intervention start-point model, a randomization statistical test is conducted on the difference between the means of all B and all A series outcomes for each of the intervention start-point divisions (or *transitions*) that could have resulted from the random-selection process (see also [Edgington & Onghena, 2007](#)). [Moreover, any other summary measure of relevance to the researcher's hypothesis about the nature of change from Phase A to Phase B (e.g., change in the series' medians, slopes, variances) can also be the focus of a randomization-test analysis.]

With the resulting set of mean differences yielding a randomization distribution, the mean difference associated with the actual intervention start point is examined to see where it falls within the set. The probability of obtaining a

IMPROVED RANDOMIZATION TESTS

mean difference as extreme as or more extreme than the actual mean difference represents the unlikelihood of the outcome. Either signed or unsigned mean differences are considered for one- and two-tailed hypothesis tests, respectively. For example, for an AB design with one case, 25 outcome-assessment periods, and 20 potential intervention start points, if the actual start point were found to produce the largest mean difference (in the predicted direction) between the B and A series outcomes, then the one-tailed significance probability associated with that event would be given by $p = 1/20 = .05$. For a two-tailed test, as or more extreme opposite-sign mean differences would also need to be taken into account. For instance, if there were a mean difference equal in magnitude but opposite in sign to the one just indicated for the actual intervention start point, then the two-tailed significance probability would be $2/20 = .10$.

In Edgington's (1975) random intervention start-point model for a one-case AB design, it is assumed that the A phase consists of a baseline series, the B phase consists of an intervention series, and that the former logically precedes the latter. With those assumptions, the number of possible outcomes (B–A mean differences) in the randomization distribution is k , the number of potential intervention start points. Accordingly, with one case, 30 total observations, and $k = 10$ potential intervention start points, if the actual B–A mean difference produced were the largest of the 10 and in the predicted direction, then the one-tailed significance probability of that outcome would be $p = 1/10 = .10$. In order to achieve statistical significance at a traditional $\alpha = .05$ level (one-tailed), one would need to include at least $k = 20$ potential intervention start points in the randomization distribution (i.e., so that if the most extreme mean difference in the predicted direction were obtained, then p would equal $1/20 = .05$). To achieve statistical significance with $\alpha = .05$ via a two-tailed test, a longer series with a minimum of $k = 40$ potential intervention start points would be required (i.e., so that $p = 2/40 = .05$ is possible).

Randomized Order (Dual Randomization) Addition to the Edgington Model

Edgington (1975) proposed his random intervention start-point design-and-analysis procedure 40 years ago. It has been incorporated into a variety of single-case intervention designs (e.g., Koehler & Levin, 1998; Levin & Wampold, 1999; Marascuilo & Busk, 1988; Onghena, 1992) and is being implemented in its original form to this day. However, it will be shown here that an addition to the procedure (referred to here as a *modified* procedure), which enhances its internal

validity by eliminating bias due AB phase-order effects, is possible and one that is applicable in a number of single-case intervention investigations. To illustrate, suppose that instead of A representing a baseline or control phase, it represents one type of experimental intervention—say, a behavioral intervention for combatting a particular phobia. In contrast, B might represent a cognitive intervention targeting the same phobia. Within that context, the case receives both interventions. To have a legitimate (unconfounded) comparison of Intervention A and Intervention B, it is imperative that the order in which the two interventions are administered to the case is randomly (rather than arbitrarily) determined. The preceding statement applies whether the investigation includes only one case or multiple cases (although in multiple-case situations, systematic counterbalancing of intervention orders across cases might be implemented to achieve the same goal).

In addition, it is worth noting that A and B need not refer only to two competing interventions. Rather, suppose that A represents a baseline, standard, or control condition and B an intervention condition. As has been suggested previously (e.g., [Kratochwill & Levin, 2010](#)), further suppose that prior to the commencement of the actual experiment, a mandatory baseline (or adaptation/warm-up) phase (A') is required of all cases. With A' included, it would then be possible, appropriate, and presumably acceptable to researchers to begin the experiment proper by randomizing each case's subsequent A and B phases (i.e., an A randomly selected to be first means that the case remains in the baseline condition, followed by the B intervention condition; and a B randomly selected to be first means that the case begins with the intervention condition, followed by the A baseline condition). Accordingly, the modified order-randomization procedure is applicable in either one- or two-intervention AB designs, with the prospect of improving both design (internal validity) and analysis (statistical-conclusion validity) of two-phase single-case intervention studies.

With intervention-order randomization built into the just-discussed one-case example based on 30 total observations and 10 potential intervention start points, in addition to the intervention start points associated with the conventional AB order of intervention administration, one would also need to consider the possibility that Intervention B had been randomly selected to be administered first. If that had happened, there would be a corresponding 10 potential intervention start points for the BA order of intervention administration, resulting in a total of $k = 20$ potential start-point outcomes that would be included in the complete randomization distribution.

Multiple-Case Extension of the Modified Edgington Model

As we will show, the order-randomization procedure applies to multiple-case (*replicated*) AB situations as well, increasing the total number of possible randomization-distribution outcomes by a factor of 2^N , where N represents the number of cases. Specifically, with N cases and one of k_i potential intervention start points randomly selected for each case, with Marascuilo and Busk's (1988) multiple-case extension of Edgington's (1975) single fixed-order intervention start-point model, a total of $\prod_{i=1}^N k_i$ randomization-distribution outcomes are possible, and in the special case for which all k_i are equal to k , this quantity reduces to k^N . With the addition of an order-randomization process to create the present dual randomization model, the total number of possible randomization-distribution outcomes increases to $\prod_{i=1}^N k_i \times 2^N$ and $k^N \times 2^N = (2k)^N$ for the general and special-case situations, respectively.

Hypothetical example

We illustrate the present random-order randomization-test procedure for a replicated single-case AB design by means of a hypothetical example. Suppose that a language researcher wishes to improve the baseline vocalization output (A phase) of two low word-producing children through some type of positive-reinforcement intervention (B phase). For the random-order version of the present example we assume that a mandatory A' baseline (warm-up) phase was initially administered, followed by a random determination of whether the first phase of the actual study would be a baseline (A) or an intervention (B) phase, thereby producing either an A'AB or A'BA design. Although in comparison to a traditional fixed-order AB design, this type of randomized AB design is more scientifically credible (especially when replicated across cases), the latter design was not considered in the current What Works Clearinghouse (WWC) single-case intervention design *Standards* (Kratochwill et al., 2010). Our hypothetical study is presented simply to illustrate both the original (Edgington, 1975) fixed-order and the present random-order randomization-test procedures, without taking into account the study's internal-validity characteristics. Consideration of internal-validity issues is included later in the Discussion section.

In this hypothetical study, the number of single-word vocalizations by each child during a 5-minute play period is recorded, with Child 1 observed in each of 25 daily sessions and Child 2 observed in each of 15 daily sessions, and where both children must be observed in at least 3 A sessions and 3 B sessions (thereby resulting in 20 and 10 potential intervention transition points for Child 1 and

Child 2, respectively). In addition, because the researcher wishes to randomize the intervention order (AB or BA) for each child, three preliminary five-minute A' warm-up sessions are provided prior to the start of the children's actual experimental sessions. An initial coin toss determines that Child 1 will be administered an AB intervention order, with the 20 potential intervention transition points specified from between the 4th and 23rd sessions inclusive and the randomly selected actual intervention transition point occurring just prior to Session 10. For Child 2, a BA intervention order results from a second coin flip, with the 10 potential intervention transitions specified from between the 4th and 13th sessions inclusive and an actual randomly selected intervention transition point just prior to the 7th observation.

The A- and B-phase observations are presented in Table 1. Given the present random-order AB intervention start-point randomization model, the data were analyzed with Gafurov and Levin's (2014) single-case *ExPRT* (*Excel® Package of Randomization Tests*) package—see Levin et al. (2014) for complete information about *ExPRT*. In Table 2 are presented the B–A mean differences associated with each of the potential intervention transition points for the two children.

The first Table 2 entry of 2.41 for Child 1, which corresponds to an A-to-B intervention transition point just prior to Observation 4, was calculated by taking the average of Child 1's Observations 4 through 25 (mean B phase = 6.41) minus the average of that child's Observations 1 through 3 (mean A phase = 4.00). The same process was followed for each of the subsequent 19 potential intervention points for Child 1, which ends with the average of that child's Observations 23 through 25 (mean B phase = 8.00) minus the average of that child's Observations 1 through 22 (mean A phase = 5.86), resulting in Child 1's final mean difference of 2.14 in Table 2. Next, and as indicated in Table 2's Footnote a, 20 additional mean differences were calculated for Child 1 under the assumption that instead of an A–B intervention order, the reverse B–A order had been selected. Under that assumption, the first mean difference for Child 1 would be $4.00 - 6.41 = -2.41$, which is exactly the same numerically but opposite in sign to the previously calculated child's first value in Table 2. The same is true for all of Child 1's calculated reverse-order values, including the 20th one, which is now -2.14 . The same process applied to Child 2's data yields the 10 actual B–A mean differences presented in Table 2 (i.e., $6.00 - 4.92 = 1.08$ for the first one), as well as 10 reverse-order and opposite-sign A–B mean differences.

IMPROVED RANDOMIZATION TESTS

Table 1. Hypothetical data for Child 1's 25-observation series, with a randomly selected AB intervention order, 20 potential intervention transition points (between Observations 4 and 23 Inclusive), and a randomly selected actual intervention transition point just prior to Observation 10; and for Child 2's 15-observation series, with a randomly selected BA intervention order, 10 potential intervention transition points (between Observations 4 and 13 Inclusive), and a randomly selected actual intervention transition point just prior to Observation 7

Child 1			Child 2		
Observation	Phase	Vocalizations	Observation	Phase	Vocalizations
1	A	4	1	B	6
2	A	3	2	B	5
3	A	5	3	B	7
4	A	5	4	B	5
5	A	2	5	B	6
6	A	5	6	B	5
7	A	3	7*	A	4
8	A	4	8	A	5
9	A	4	9	A	3
10*	B	5	10	A	5
11	B	6	11	A	4
12	B	7	12	A	5
13	B	6	13	A	6
14	B	7	14	A	5
15	B	8	15	A	6
16	B	7			
17	B	9			
18	B	8			
19	B	6			
20	B	8			
21	B	9			
22	B	8			
23	B	7			
24	B	9			
25	B	8			

*Actual intervention transition point.

Table 2. The B–A mean difference associated with: (1) each of Child 1’s 20 potential intervention transition points (O_4 – O_{23}) for a randomly selected AB intervention order; and (2) each of Child 2’s 10 potential intervention transition points (O_4 – O_{13}) for a randomly selected BA intervention order

Potential Intervention Point	Child 1 B-A Mean Difference ^a	Child 2 B-A Mean Difference ^b
O_4	2.41	1.08
O_5	2.23	0.84
O_6	2.90	1.00
O_7	2.79	0.89*
O_8	3.14	0.55
O_9	3.30	0.52
O_{10}	3.49*	-0.06
O_{11}	3.53	-0.10
O_{12}	3.46	-0.50
O_{13}	3.28	-0.67
O_{14}	3.29	
O_{15}	3.19	
O_{16}	2.97	
O_{17}	2.94	
O_{18}	2.58	
O_{19}	2.41	
O_{20}	2.69	
O_{21}	2.60	
O_{22}	2.24	
O_{23}	2.14	

*Mean difference associated with the actual intervention transition point. ^a The 20 A–B mean differences are also calculated and added to these to form a 40-outcome randomization distribution; all of the A–B mean differences are the same as the corresponding B–A mean differences given here but opposite in sign. ^b The 10 A–B mean differences are also calculated and added to these to form a 20-outcome randomization distribution; all of the mean A–B differences are the same as the corresponding mean B–A differences given here but opposite in sign.

The resulting joint randomization distribution therefore contains 40 mean differences for Child 1 combined with 20 mean differences for Child 2, for a total of $40 \times 20 = 800$ averaged mean differences (i.e., Child 1’s 1st mean difference averaged with Child 2’s 1st mean difference, Child 1’s 1st mean difference averaged with Child 2’s 2nd mean difference, all the way up to and including Child 1’s 40th mean difference averaged with Child 2’s 20th mean difference). When that is done by the *ExPRT* program, it is found that the *actual* joint mean

IMPROVED RANDOMIZATION TESTS

difference that was obtained in the study is 2.19, which is Child 1's mean difference associated with that child's actual intervention transition point of O_{10} (3.49) averaged with Child 2's actual intervention transition-point mean difference of O_7 (.89). Of the 800 outcomes in the joint randomization distribution, a value of 2.19 is the 10th highest, which results in a one-tailed significance probability of $p = 10/800 = .0125$. For this example, had a one-tailed Type I error probability (α) of .05 been selected, it could be concluded that the positive-reinforcement intervention (B) distribution values differed statistically from those in the baseline distribution (A), with the additional inference that the former distribution's values were higher. We note that both here and in the various simulations conducted in the present series of investigations, one-tailed tests are conducted because it is assumed that [especially in single-case A (baseline) – B (intervention) research] the researcher has a clear and defensible rationale for the direction of change that is associated with the intervention.

Insofar as randomization tests are not tailored to test for the equality of two populations' specific parameters, all that can be tested for is the equality of the two population distributions *per se*. For the present randomization test, the test statistic involves sample-mean differences and because that is the test that produced a statistically significant result here (favoring the intervention phase over the baseline phase), a reasonable inference is that there was an A- to B-phase upward shift in the children's level of responding.

Advantages of the Order Randomization Modification

The present order-randomization approach enhances the internal validity of a single-case AB design by virtue of its removing bias stemming from intervention-order effects. As an important byproduct, the approach also elevates the status of the basic AB single-case intervention design from a WWC *Standards* "acceptable design" standpoint (Kratochwill et al., 2010), particularly when replicated across independent participants at different points in time. According to the WWC *Standards*, two-phase A (Baseline) – B (Intervention) designs are not scientifically credible (and therefore unacceptable) because they suffer from too many potential sources of internal invalidity. For extended discussion of acceptable designs, see Kratochwill, et al. (2010, 2013).

Including outcomes from both intervention-administration orders in the randomization distribution also provides fundamental pragmatic advantages for single-case intervention researchers. First, with the original Edgington (1975) model, a researcher would need to designate 20 potential intervention start points

(based on at least 21 total observations) to produce a randomization test that is capable of detecting an intervention effect with a one-tailed Type I error probability less than or equal to .05. With the present procedure, a researcher would need to designate only half as many potential intervention start points (here, 10, based on a total of 11 total observations, resulting in 20 possible outcomes) to detect an intervention effect. A related reason why the present procedure has practical importance for single-case intervention researchers is that (and as will be demonstrated here) relative to the original Edgington (1975) model, the modified approach may produce statistical-power advantages as well. Thus, for no more expense than a coin to flip, a researcher might reap both methodological and statistical benefits by adopting the present dual-randomization procedure rather than either the original single-randomization Edgington model or Marascuilo and Busk's (1988) multiple-case extension of it.

Relationship to Traditional Experimental Designs and Statistical Analyses

Although unrecognized at the time that the present order-randomization approach was initially conceptualized, its logic maps directly onto a statistical procedure in the traditional group randomized treatment-design literature. In particular, consider a randomized two-treatment correlated-samples (or within-subjects) design based on N participants, to which a nonparametric randomization test is applied as an appropriate alternative in (especially small-sample) situations where the normality assumption of a correlated-samples t test (or a one-sample repeated-measures analysis) is questionable.

To illustrate that situation, we revisit an example that was recently presented by Ferron and Levin (2014, p. 174). Suppose that in a sample of $N = 8$ adults, each participant is administered two different fear-reducing treatments, A (a behavioral treatment) and B (a cognitive intervention), with the former posited to be more effective than the latter. It is determined in advance that the equal-effectiveness hypothesis will be tested with a randomization test based on a one-tailed α of .05. To produce a scientifically credible experiment, the order in which the two treatments are administered is again randomly determined on a case-by-case basis by means of coin flips: say, heads represents an AB order and tails a BA order. On the basis of that process, let us suppose that 5 participants ended up in the AB condition and 3 in the BA condition. Following the administration of each treatment, participants' fear responses are assessed on a 7-point Likert scale, with higher numbers indicating greater fear. With the measure of interest defined

IMPROVED RANDOMIZATION TESTS

as the difference between each participant's B and A ratings (i.e., $B-A$), the following outcomes were obtained for the 8 participants:

$$+3.0 \quad +3.5 \quad -1.5 \quad +2.0 \quad +4.5 \quad +3.5 \quad -2.0 \quad +4.0$$

The observed test statistic is given by the average of these differences, which is equal to $+17/8 = 2.125$. A randomization distribution is created from the $2^N = 2^8 = 256$ possible ways in each + and - signs could be attached to these 8 numerical values. For example, the first outcome in the randomization distribution (with all + signs) would be:

$$+3.0 \quad +3.5 \quad +1.5 \quad +2.0 \quad +4.5 \quad +3.5 \quad +2.0 \quad +4.0$$

yielding a mean difference of $+24/8 = 3.000$, and the last (with all minus signs) would be:

$$-3.0 \quad -3.5 \quad -1.5 \quad -2.0 \quad -4.5 \quad -3.5 \quad -2.0 \quad -4.0$$

yielding a mean difference of $-24/8 = -3.000$. The remaining 254 possible outcomes would fall somewhere between these two extremes.

The actually obtained mean difference of $+2.125$ appears to be on the higher side of this distribution. In fact, it turns out to be among the 9 highest possible outcomes (specifically, an outcome that is exceeded by only 5 outcomes and that is tied with 3 others). Accordingly, a one-tailed test of the hypothesis that the A and B treatments have equal distributions would be associated with a p -value (consistent with the alternative hypothesis that Treatment B is producing higher fear ratings than Treatment A) that is equal to $9/256 = .035$. Because this value is less than the predetermined α of .05, it would be concluded that the actually obtained mean difference of $+2.125$ is statistically significant.

Note that for this conventional-group design and associated randomization test, the all-possible assignment of + and - signs to the 8 absolute $B-A$ differences corresponds exactly to the logic and operationalization of the single-case AB order-randomization procedure to be investigated here. In particular, the procedure incorporates two separate forms of randomization for each of the N participating cases, Edgington's intervention start-point randomization and AB order randomization. In the simplest situation where there is only one potential intervention start point for each case (as in the just-presented $N = 8$ example), the total number of possible start-point randomizations is equal to $k^N = 1^8 = 1$. The

present order-randomization procedure involves each of the 8 participants contributing two differences (i.e., B–A and A–B) to the randomization distribution, resulting in $2^N = 2^8 = 256$ joint randomization outcomes, and which, according to the previously given special-case dual-randomization formula, $k^N \times 2^N$, yields a total of $1 \times 2^8 = 256$ possible randomization outcomes. This total is identical to the number of possible randomization-distribution outcomes associated with the just-presented example. It is instructive to note that the total number of possible randomization outcomes associated with order randomization can be alternatively expressed as $\sum_{x=0}^N \binom{N}{x}$, where N = the number of cases and x = the number of positive B–A differences that could be associated with the N actual outcomes. For the present example, this expression is equal to $\sum_{x=0}^8 \binom{8}{x}$, or

$$\begin{aligned} & \binom{8}{0} + \binom{8}{1} + \binom{8}{2} + \binom{8}{3} + \binom{8}{4} + \binom{8}{5} + \binom{8}{6} + \binom{8}{7} + \binom{8}{8} \\ &= 1 + 8 + 28 + 56 + 70 + 56 + 28 + 8 + 1 \\ &= 256 \end{aligned}$$

Thus, when there is only one potential intervention point for each case and the AB design includes multiple observations, the present randomized-order test based on the difference between the A- and B-phase means maintains the same correspondence with a conventional-group correlated-samples randomization test as was shown here. Implicit in the conventional correlated-samples test is that with random assignment to treatment conditions, outcomes representing both orders of treatment administration need to be considered in the randomization test distribution. As such, the present order-randomization procedure is not really a special case at all, but rather the single-case analog of a correlated-samples randomization t test.

Focus of the Present Investigations

The focus of our series of simulation investigations was to examine the Type I error and statistical power characteristics of the dual-randomization modification (intervention start-point *plus* intervention order) relative to those of Edgington's (1975) and Marascuilo and Busk's (1988) original single-randomization (intervention start-point) test procedures. In this study we present randomized intervention-order findings not just for a basic two-phase AB design, but also for a randomized pairs variation of that design (Levin & Wampold, 1999), a single-

case adaptation of the conventional-group crossover design, and Onghena's (1992) four-phase ABAB design.

Investigations 1-3: Randomized Intervention Order for the Basic AB Design

Investigation 1

Method In Investigation 1, the focus was on 30-observation designs for a single participant (i.e., $N = 1$), where the intervention start point was randomly selected from the middle 20 observations. The series length of 30 was chosen for initial examination because: (1) 20 start points is the minimum number needed to obtain a statistically significant result with a one-tailed α of .05 for an AB randomized start-point design with one case; and (2) the WWC *Standards* require a minimum of five observations in each phase (Kratochwill et al., 2010, 2013).

Data were generated using SAS IML (SAS, 2013), where the time-series data were obtained by adding an error vector to an effect vector. The error vector was created such that it was distributed normally and had an autocorrelation of 0 or .3 by using SAS's autoregressive moving-average simulation function (ARMASIM). The autocorrelation values of 0 and .3 were motivated by a survey of actual single-case studies where it was reported that the average autocorrelation was .2, after adjusting for bias in the estimates (Shadish & Sullivan, 2011). To obtain simulated errors based on an autocorrelation of .3, the autoregressive parameter matrix was set to $\{1 \ - .3\}$, the moving average parameter matrix was set to $\{1 \ 0\}$, and a standard deviation of the independent portion of the error was set to 1.0 (for details on the simulation algorithm see Woodfield, 1988). The effect vector was coded to have values of 0 for all baseline observations, and values of d for all intervention phase observations, and thus d corresponds to the mean shift between intervention and baseline observations in standard deviation units, $(\mu_B - \mu_A)/\sigma$ (see Busk & Serlin, 2005), where the standard deviation is based on the independent portion of the within-case error term (see, for example, Levin, Ferron, & Kratochwill, 2012) (for an alternative operationalization of d that corresponds mathematically to a conventional groups effect-size measure, see Shadish et al. (2014)). The value of d was varied to examine the one-tailed Type I error probability for $d = 0$ and the powers for d s ranging from .5 to 5 in increments of .5. For reference, if the d used for the present data generation is estimated for each of the 200 Phase A-to-Phase B contrasts examined in the survey of single-case interventions reported by Parker and Vannest (2009), the empirically

observed values of d (assuming no autocorrelation for simplicity) for the 10th, 50th, and 90th percentile ranks are estimated to be 0.46, 1.70, and 3.88, respectively.

By crossing each design (single, dual), with each level of autocorrelation ($r = 0, .3$), and each effect size ($d = 0$ to 5, in increments of .5), $2 \times 2 \times 11 = 44$ conditions were obtained, and for each of these conditions the data for 10,000 studies were simulated. The data for each simulated data set were analyzed using a randomization test in which the obtained test statistic ($M_B - M_A$) was compared to the complete randomization distribution. The proportion of simulated studies in which the randomization test led to a one-tailed p -value of .05 or less was determined to estimate the rejection rate (Type I error or power) of the randomization test for each of the 44 experimental conditions.

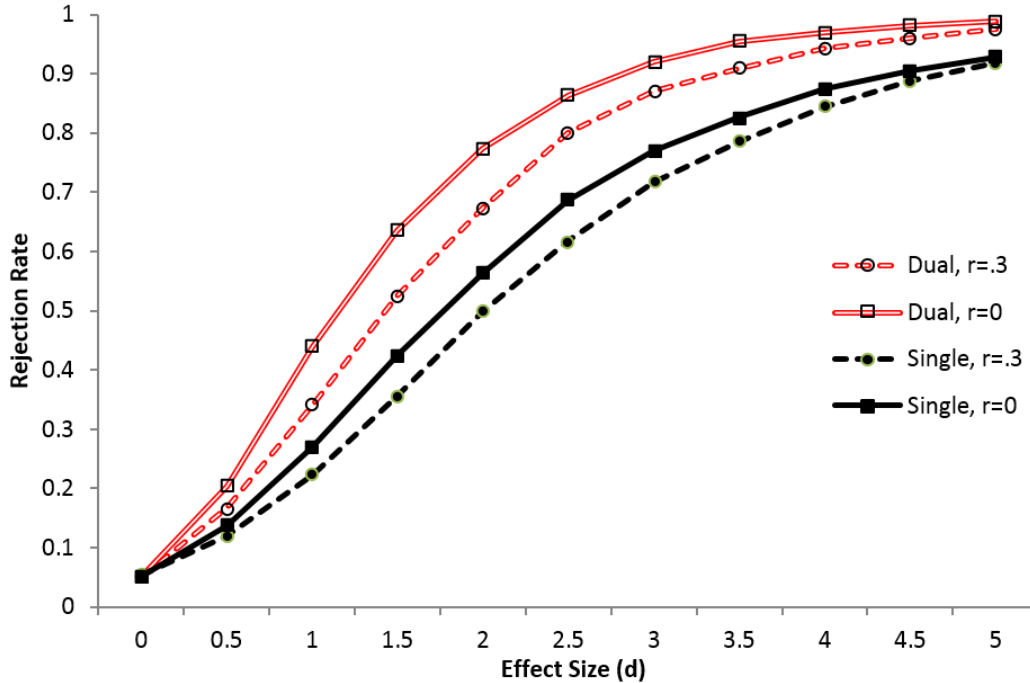


Figure 1. Investigation 1: Comparison ($\alpha = .05$, one-tailed) of randomization tests for a one-case ($N = 1$) AB randomized intervention start-point design (Single) and the randomized intervention start-point plus randomized intervention- order design (Dual), where the start point was randomly selected between the 6th through the 25th observations inclusive in a 30-observations study. The rejection rate of the null hypothesis is shown as a function of the effect size and level of autocorrelation.

IMPROVED RANDOMIZATION TESTS

Results Results are shown in Figure 1 for Edgington's (1975) original procedure (single) and for the present randomized-order modification (dual). As may be seen in that figure, when the effect size is 0, all situations are associated with empirical powers (which, for $d=0$ are equivalent to Type I error probabilities) that correspond to their nominal .05 values. Not surprisingly, based on previous findings (e.g., Ferron & Sentovich, 2002; Ferron & Ware, 1995; Levin et al., 2011), it may also be seen that for $ds > 0$ power is uniformly higher for $r=0$ than for $r=.3$. As the effect size increases, so does power, although more rapidly for the dual-randomization procedure than for its single-randomization counterpart. The largest power differences, favoring the former, reach .21 in the $r=0$ situation for ds of 1.5 and 2.0; and in the $r=.3$ situation the largest power difference is .18 for a d of 2.5.

Investigation 2

Method In Investigation 2, series length (i.e., the number of observations) was systematically varied for a single-participant ($N=1$) design, while holding the effect size constant at $d=2$. A d of 2 was chosen because it is a large enough effect to typically be of interest to a single-case researcher. Yet, a d of 2 is small enough that it is not readily detectable (power $< .80$) in a single-participant 30-observations design when there is a moderate autocorrelation of .30 and applying either the single- or dual-randomization approach (as may be seen in Figure 1, where powers are .50 and .67, respectively). The simulation methods paralleled those of the initial investigation (including a one-tailed α of .05), but d was held constant at 2.0 for all conditions and series length was varied from 20 to 150 in increments of 10. The number of potential intervention start points was always the series length minus 10 to ensure at least five observations in the baseline and intervention phases.

Results Results for this set of simulations are provided in Figure 2, where with an autocorrelation of .30, power of at least .80 is attained for the dual-randomization approach with 60 observations (power = .81), in contrast to the single-randomization design where .80 power is not quite attained even with 150 observations (power = .79). For 30 to 100 observations, the power difference between the two randomization schemes (favoring dual) ranges from .13 to .31 when the autocorrelation is 0 and from .17 to .30 when the autocorrelation is .30.

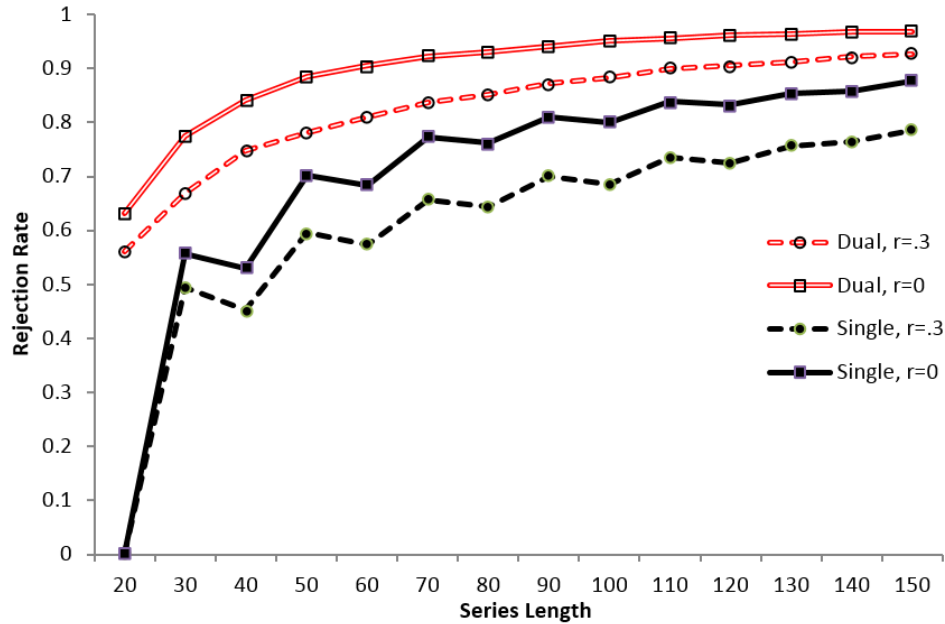


Figure 2. Investigation 2: Comparison ($\alpha = .05$, one-tailed) of randomization tests for a one-case ($N = 1$) AB randomized intervention start-point design (Single) and the randomized intervention start-point plus randomized intervention-order design (Dual). The rejection rate of the null hypothesis is shown as a function of series length and level of autocorrelation. The effect size is 2.0 and the number of potential intervention start points (x) is equal to the series length minus 10 and encompasses the middle x observations.

It should be noted that the power is 0 for the single-randomization scheme with 20 observations because there are only 10 possible intervention start points and thus statistical significance cannot be obtained at the one-tailed .05 level. In addition, the undulation in the power curves for the single-randomization approach makes sense when one recognizes that: (1) for a series length of 30, statistical significance with $\alpha = .05$ can be attained only for the most extreme of the 20 permutations; and (2) with a series length of 40, statistical significance can again be attained only for the most extreme permutation, but now there are 30 permutations and so the most extreme is somewhat more difficult to achieve. Although power drops for the 40-observation series, with a series length of 50, statistical significance can be attained for either of the two most extreme permutations and thus power jumps back up again.

IMPROVED RANDOMIZATION TESTS

Investigation 3a

Method In Investigation 3a, the effect of multiple-case replications (i.e., $N > 1$) on the power of the single- and dual-randomization procedures was examined. More specifically, a design with 15 observations and 5 potential intervention start points, randomly selected from observations 6 through 10, was examined with 2, 3, 4, 5, and 6 participants based on a one-tailed α of .05. For the single-randomization approach, 7 and 8 participants were also included. These numbers of participants seemed reasonable given the survey by Shadish and Sullivan (2011), in which it was found that the number of cases in single-case studies averaged 3.64, with a range of 1 to 13. In the present study, effect sizes varied from 0 to 3 in increments of .5 and the autocorrelation was set either to 0 or .3.

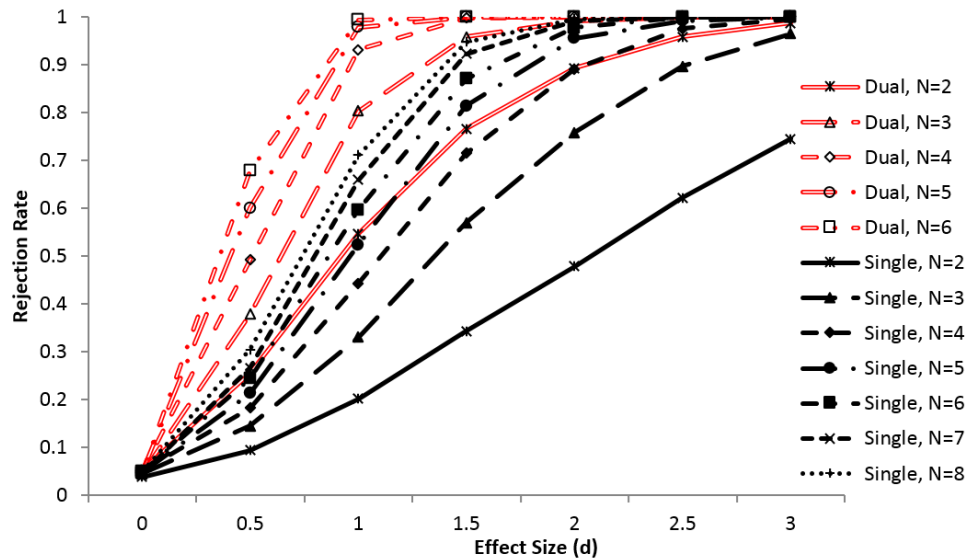


Figure 3. Investigation 3a: Comparison ($\alpha = .05$, one-tailed) of randomization tests for the Single and Dual basic AB randomized designs replicated across N cases. The rejection rate of the null hypothesis is shown as a function of effect size and N , for a 15-observations design with 5 potential intervention start points designated from between the 6th and 10th observations inclusive and an autocorrelation of 0.

Results Results from simulations where the autocorrelation is 0 are shown in Figure 3, whereas those for an autocorrelation of .3 are shown in Figure 4. In both figures, it may be seen that for all sample sizes the empirical Type I error probabilities are well controlled at .05 for both the single- and dual-randomization approaches. The important thing to note is that in both figures, for all effect sizes the dual approach based on as few as $N = 3$ participants has associated power that is greater than or equivalent to the single approach based on $N = 8$ participants. For example, in Figure 4 it may be seen that with an autocorrelation of .3, $N = 3$ dual- and $N = 8$ single-randomization powers are .66 and .61, respectively, for an effect size of 1.0; and they are .90 and .89, respectively, for an effect size of 1.5.

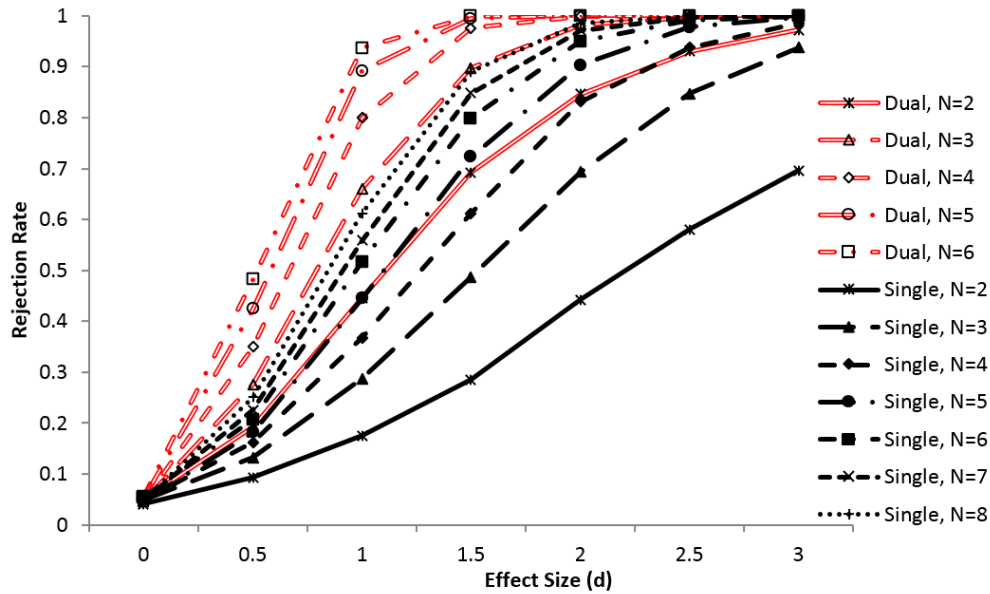


Figure 4. Investigation 3a: Comparison ($\alpha = .05$, one-tailed) of randomization tests for the Single and Dual basic AB randomized designs replicated across N cases. The rejection rate of the null hypothesis is shown as a function of effect size and N , for a 15 observations design with 5 potential intervention start points designated from between the 6th and 10th observations inclusive and an autocorrelation of .3.

Investigation 3b

Method In this investigation, the simulations of Investigation 3a were replicated with the sole difference being that a two-tailed test with $\alpha = .05$ was conducted, as opposed to a one-tailed test.

Results The results are summarized in Figure 5 for an autocorrelation of 0 and in Figure 6 for an autocorrelation of .3. Again, it may be seen that all of the empirical Type I errors are at the expected .05 level for both autocorrelation values. Although the Investigation 3a results (i.e., the equivalence of dual-randomization $N = 3$ and single-randomization $N = 8$) were not identical here, the general pattern was. In this case, however, the appropriate power equivalence involves dual $N = 4$ and single $N = 8$. Specifically, in Figure 6 it may be seen that with an autocorrelation of .3, the former and latter powers are .65 and .61, respectively, for an effect size of 1.0; and they are .93 and .89, respectively, for an effect size of 1.5.

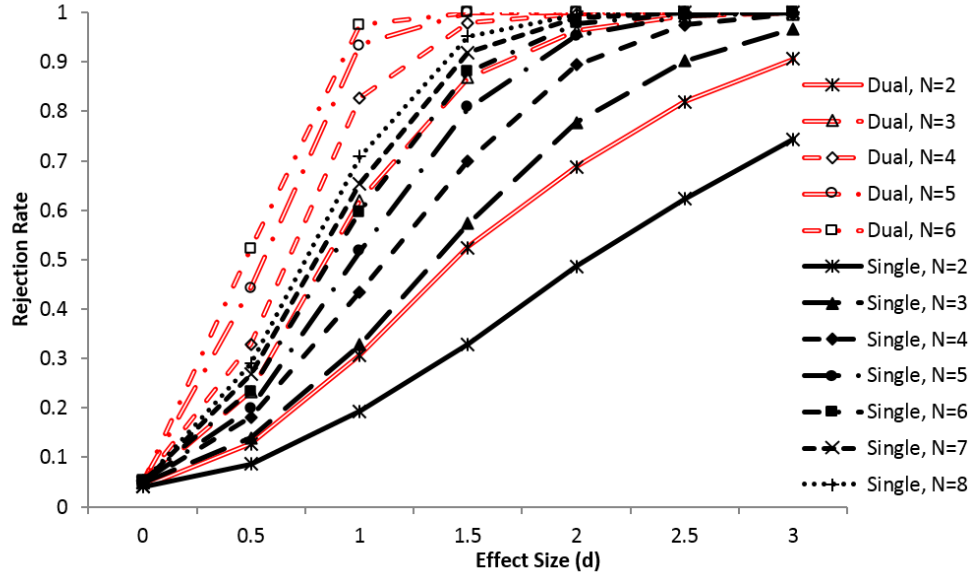


Figure 5. Investigation 3b: Comparison ($\alpha = .05$, two-tailed) of randomization tests for the Single and Dual basic AB randomized designs replicated across N cases. The rejection rate of the null hypothesis is shown as a function of effect size and N , for a 15 observations design with 5 potential intervention start points designated from between the 6th and 10th observations inclusive and an autocorrelation of 0.

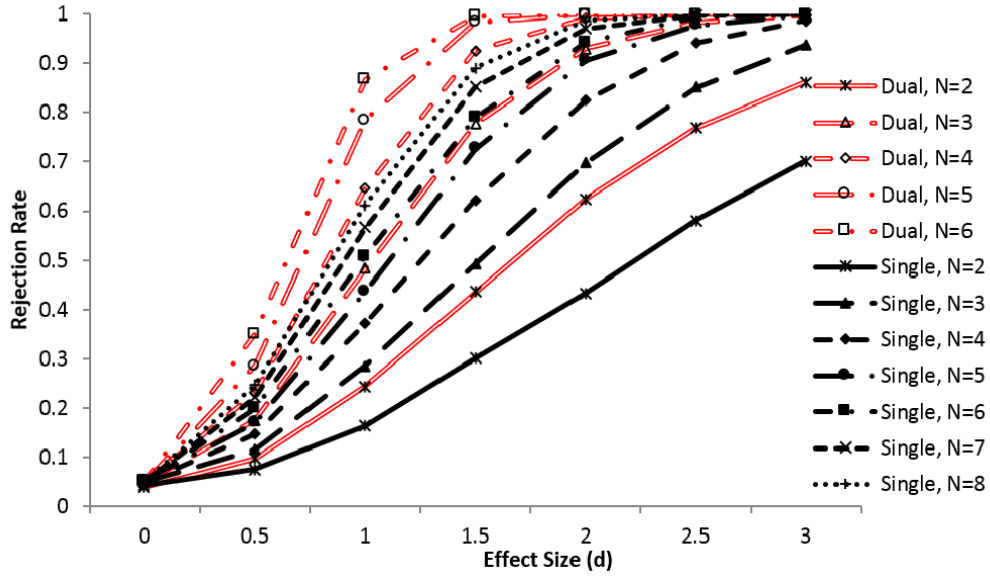


Figure 6. Investigation 3b: Comparison ($\alpha = .05$, two-tailed) of randomization tests for the Single and Dual basic AB randomized designs replicated across N cases. The rejection rate of the null hypothesis is shown as a function of effect size and N , for a 15 observations design with 5 potential intervention start points designated from between the 6th and 10th observations inclusive and an autocorrelation of .3.

Thus, in the present investigation we observe that for two-tailed tests the dual-randomization power benefits (relative to single randomization) are comparable to those reported for Investigation 3a's one-tailed tests. It is important to point out, however, that the situations examined here were all based on multiple-case ($N > 1$) designs. It turns out that for the special-case $N = 1$ situation, although the dual- over single-randomization power advantage is evident when one-tailed tests are conducted (as was true in Investigations 1 and 2), the dual- and single-randomization schemes yield equivalent power results with two-tailed tests. Because the two-tailed test is based on randomization-distribution absolute-value outcomes, the dual-randomization distribution contains every outcome of the single-randomization distribution as well as its opposite-order complementary outcome, thereby yielding exactly the same p-value for each test. (To illustrate these notions, see Child 1's hypothetical data, including Footnote a in Table 2. The 40 unsigned mean differences (i.e., 20 $|B-A|$ plus 20 $|A-B|$) would constitute the dual-randomization distribution for a two-tailed test). Because there are across-case combinations when $N > 1$, there is no longer a one-to-one

correspondence between the single- and dual-randomization distributions and so their powers will generally differ, with the latter being greater (as was observed in Figures 5 and 6).

Investigation 4: Randomized Intervention Order and/or Randomized Intervention Assignment in Levin and Wampold's (1999) AB Pairs Design

Another type of dual-randomization strategy is possible when a case consists of a pair of participants, as in Levin and Wampold's (1999) simultaneous intervention start-point model. With the Levin-Wampold model, N participant (or other unit) pairs are created and the members of each pair are randomly assigned to two different intervention conditions (or to an intervention and control condition), X and Y. With this model, Levin and Wampold presented two hypotheses that would be of interest to researchers: (1) a *general intervention effectiveness* hypothesis, namely that averaged across the two intervention conditions, there is no difference between Phase A and Phase B performance (analogous to the time main effect in a conventional two-treatment pretest-posttest design); and (2) a *comparative intervention effectiveness* hypothesis, namely that the change in participants' performance from Phase A to Phase B is the same in the two intervention conditions (analogous to the treatment-by-time interaction in a conventional two-treatment pretest-posttest design). Unrecognized by Levin and Wampold at the time, the randomization test of each of these hypotheses could potentially benefit from an additional randomization component. For the general intervention effectiveness hypothesis, that component is AB order randomization of the kind that we have considered in Investigations 1-3, either with or without a mandatory A' baseline phase; and for the comparative intervention hypothesis, that component consists of within-pair intervention randomization, wherein pair members are randomly assigned to the two intervention conditions.

Implementing either of these randomization types increases the total number of possible outcomes from $\prod_{i=1}^N k_i$ for Levin and Wampold's (1999) original single randomization-test procedure (i.e., the number of potential intervention start points for each pair) to $2^N \times \prod_{i=1}^N k_i$ for the present dual approach (i.e., either the number of possible random assignments of AB orders or the number of possible random assignments of interventions to pair members, times the number of potential intervention start points for each pair). In Investigation 4, we examine the statistical power consequences associated with the dual approach's additional

randomization component, for both the general and the comparative intervention effectiveness hypotheses.

Method

A power comparison of dual versus single randomization for the two hypotheses (general and comparative intervention effectiveness) was conducted with a one-tailed α of .05. Specifically, designs with 2, 3, and 4 pairs of participants were examined based on 15 observations per participant. There were 5 potential start points for each pair, randomly selected from observations 6 through 10. For the general intervention effectiveness simulations, with single randomization each pair received the baseline phase (A) followed by the intervention (B) phase; in contrast, with dual randomization the pairs were randomly assigned to either an AB or BA order. For the comparative intervention effectiveness simulations, with single randomization the first pair member always received Intervention X and the second pair member Intervention Y; in contrast, with dual randomization, pair members were randomly assigned to the two intervention conditions.

The time-series data for each case were simulated as described in the previous investigations, with the standardized effect size for the pair member assigned to Intervention X set to d_1 and the standardized effect for the pair member assigned to Intervention Y set to d_2 . For the general intervention effectiveness test, $d = (d_1 + d_2)/2$ was varied from 0 to 3 in increments of .5 by setting $d_1 = d_2 = d$. For the comparative intervention effectiveness test, $d = d_2 - d_1$, d_1 was set to 0 and d_2 was varied from 0 to 3 in increments of .5. The latter effect size can be alternatively written as $d = [(\mu_{B2} - \mu_{A2}) - (\mu_{B1} - \mu_{A1})]/\sigma$, which is readily conceptualized and interpreted as a standardized ‘difference in differences’ (e.g., Marascuilo & Levin, 1970). The present measure differs from the standardized ‘half difference in differences’ effect-size estimator of $(d_2 - d_1)/2$ that is provided in Gafurov and Levin’s (2014) *ExPRT* program for the comparative intervention effectiveness hypothesis. The half difference-in-differences measure was incorporated into *ExPRT* because it represents a properly scaled interaction contrast when formulated for sample-size and power determination purposes from an analysis-of-variance perspective (Levin, 1997). It therefore should be kept in mind that a present power estimate associated with a difference-in-differences effect size of 2.00 corresponds to the power estimate associated with *ExPRT*’s half difference-in-differences effect size of 1.00.

Results

General intervention effectiveness hypothesis Dual- and single-randomization powers for Levin and Wampold's (1999) general intervention effectiveness hypothesis are presented in Figures 7 and 8 for autocorrelations of 0 and .3, respectively. The averaged pair power results presented in Figures 7 and 8 are easy to describe, especially when juxtaposed with Investigation 3a's individual results that were previously presented in Figures 3 and 4. Although the actual power values differ in the two investigations, the patterns involving single- and dual-randomization powers—namely, the magnitudes of the power advantage favoring the latter over the former—are remarkably similar. For example, when the total number of cases is held constant (e.g., 4 individuals in Investigation 3a, 2 pairs here; 6 individuals in Investigation 3a, 3 pairs here), with an autocorrelation of .3, mid-range effect-size values of $d = 1$ and 1.5, and two asymptotic power situations excluded, the six differences between the dual- and single-randomization powers all hover around .40. Specifically, from the graphs based on $N = 4$ individuals (Figure 4) and $N = 2$ pairs (Figure 8), it may be determined that the respective power differences are .43 and .37 for $d = 1$ and are .36 and .39 for $d = 1.5$; for $N = 6$ individuals and $N = 3$ pairs, the power differences are .42 and .40 for $d = 1$.

Comparative intervention effectiveness hypothesis Dual- and single-randomization powers associated with Levin and Wampold's (1999) comparative intervention effectiveness hypothesis are presented in Figures 9 and 10 for autocorrelations of 0 and .3, respectively. In each of those figures it may be seen that the dual-randomization procedure, which incorporates additional randomization-distribution outcomes as a result of randomly assigning pair members to the two interventions, X and Y, produces substantial power increases over Edgington's (1975) original single-intervention start-point procedure. For example, in Figure 10 based on an autocorrelation of .3, $N = 3$ pairs, and a difference-in-differences effect size of 2.0 (which corresponds to *ExPRT*'s half difference in differences of 1.0), power for the dual-randomization procedure is .87 as compared to only .46 for the single-randomization procedure.

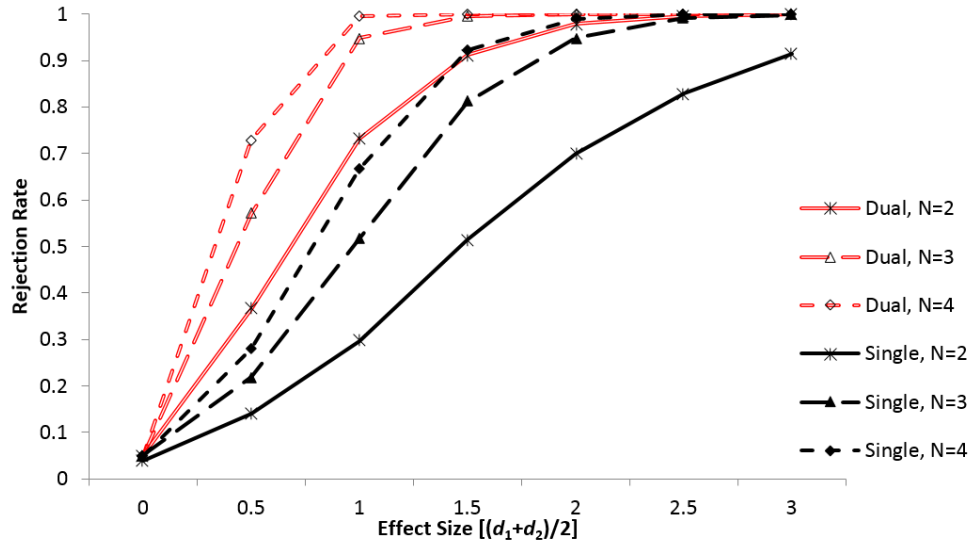


Figure 7

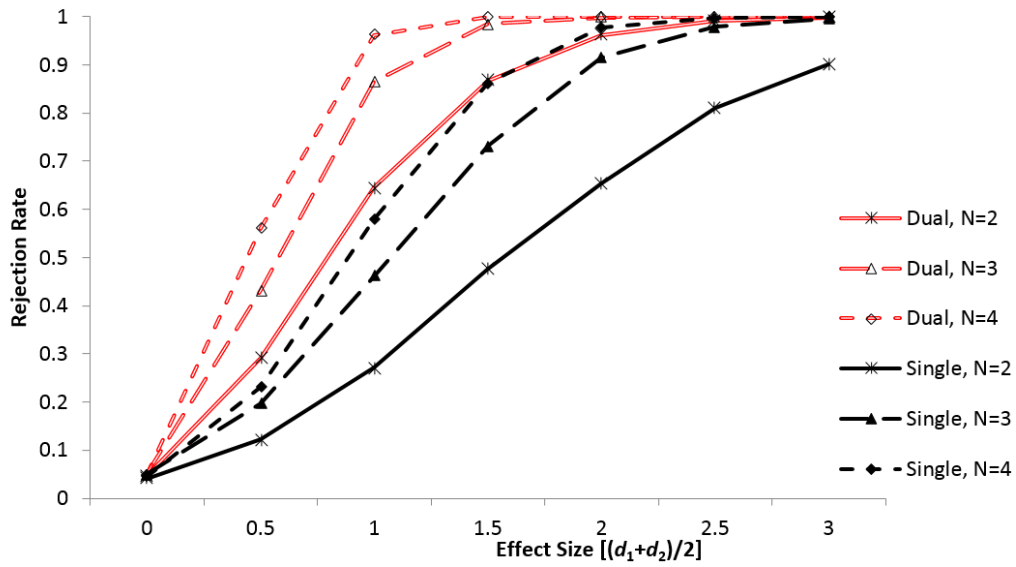


Figure 8

Figures 7 and 8. Investigation 4: Comparison ($\alpha = .05$, one-tailed) of powers for the Single and Dual randomized general intervention effectiveness hypothesis replicated across N pairs. The rejection rate of the null hypothesis is shown as a function of effect size and N , for a 15 observations design with 5 potential intervention start points designated from between the 6th and 10th observations inclusive and an autocorrelation of 0 (Figure 7) or .3 (Figure 8).

IMPROVED RANDOMIZATION TESTS

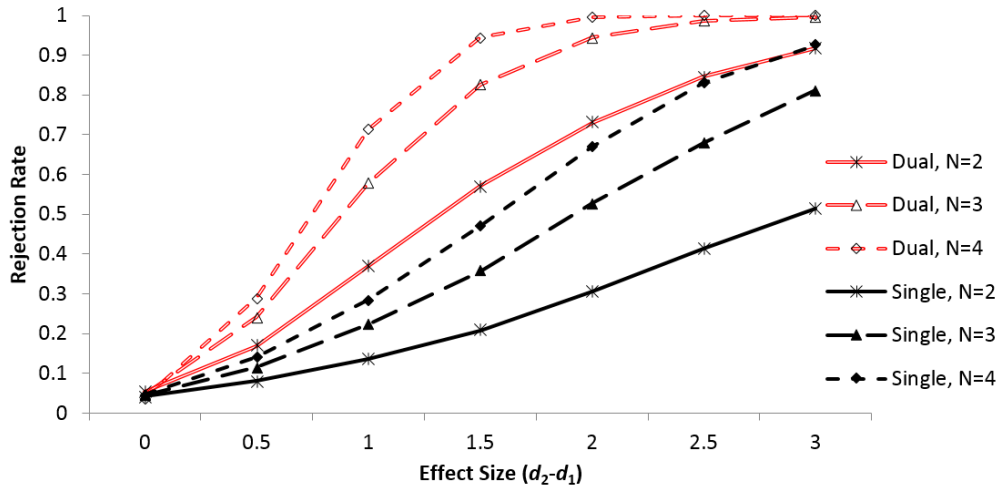


Figure 9

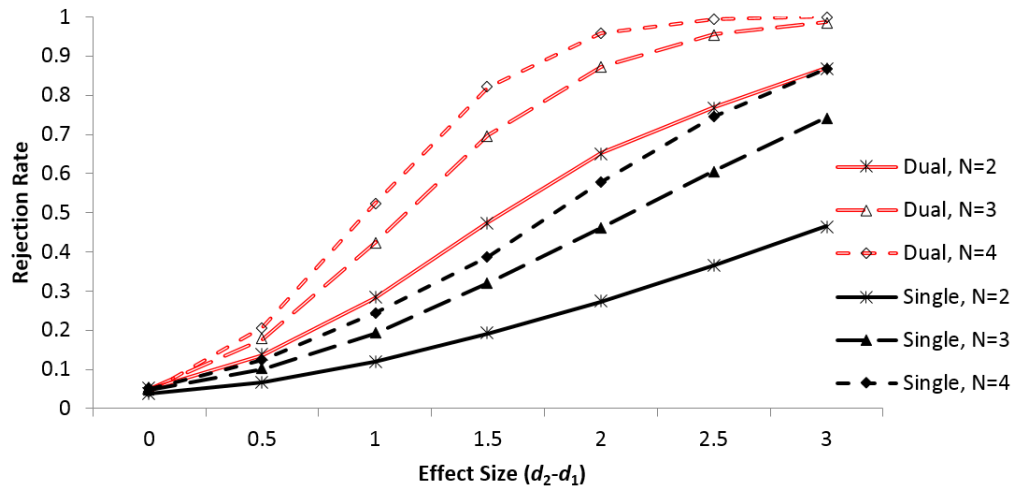


Figure 10

Figures 9 and 10. Investigation 4: Comparison ($\alpha = .05$, one-tailed) of powers for the Single and Dual randomized Levin-Wampold comparative intervention effectiveness hypothesis replicated across N pairs. The rejection rate of the null hypothesis is shown as a function of effect size and N , for a 15 observations design with 5 potential intervention start points designated from between the 6th and 10th observations inclusive and an autocorrelation of 0 (**Figure 9**) or .3 (**Figure 10**). Effect sizes are defined in a difference-in-differences metric, which correspond to half difference-in-differences effect sizes given by the present values divided by 2 (see text for further discussion).

Discussion

The present single-randomization powers associated with both the general and comparative intervention effectiveness hypotheses are quite consistent with those reported in an earlier simulation study by Lall and Levin (2004). However, the results of Investigation 4 make it clear that whenever either AB phase randomization is employed (general intervention effectiveness hypothesis, as is also manifested in Investigations 1-3) or the pair members are randomly assigned to the two intervention conditions, X and Y (comparative intervention effectiveness hypothesis), then the researcher can justifiably incorporate that randomization component into the randomization test. Doing so produces a large power boost relative to Levin and Wampold's (1999) original randomization tests that incorporate only intervention start-point randomization. The impressive dual-randomization power increases for the comparative intervention hypothesis are particularly noteworthy and heretofore undocumented. Although Levin and Wampold recognized the methodological (internal validity) necessity of randomly assigning the XY pair members to intervention conditions when testing that hypothesis, their single-randomization test procedure does not capitalize on the statistical power benefits that result from random assignment.

At the same time, and as was suggested by Levin and Wampold (1999, p. 78), now suppose that instead of X and Y representing two alternative interventions to which pair members are randomly assigned (as was examined here), they represent some non-randomly assigned participant-differentiating (or *status*) variable of interest (e.g., gender, age, ability, amount of prior experience), where one pair member (X) represents one level of the status variable (e.g., male, older, higher, more prior experience) and the other pair member (Y) represents a different level (female, younger, lower, less prior experience). In that nonrandom-assignment situation, the additional 2^N X vs. Y randomization outcomes of the modified Levin-Wampold formula (provided earlier in this section) cannot be incorporated into the randomization distribution, in which case the statistical test would revert to the original procedure developed by Levin and Wampold. It should be noted, however, that: (1) the inclusion of the status variable (e.g., gender, age, ability, amount of prior experience) still permits the investigation of a possible intervention-by-status interaction (e.g., the intervention is relatively more effective for individuals with less prior experience than for individuals with more prior experience) with the comparative intervention effectiveness test; and (2) if AB phase randomization is included in a nonrandomized status-variable study, then the 2^N factor associated with phase randomization in the modified Levin-Wampold general intervention effectiveness formula (provided earlier in

IMPROVED RANDOMIZATION TESTS

this section and the primary focus of the present order-randomization study) would reappear.

Let us additionally consider a participant-pairs situation in which both the XY and the AB factors have randomized components. For example, X and Y could represent two randomly assigned instructional interventions: experimental vs. control (as in Investigation 4, and the primary factor of interest); and A and B could represent two types of practice: teacher- vs. self-directed (the secondary factor of interest), the order of which is randomly assigned to each pair. In that situation, the currently investigated two-factor randomization design (intervention start points and phase orders) could be expanded to encompass a third randomized factor (intervention start points, instructional intervention, and practice-type phase order). Yet, it is important to note that: (1) incorporating *either* AB or XY randomization into the Levin-Wampold (1999) simultaneous pairs design will enhance the design's internal validity and produce a statistical power increase to detect general (AB) or comparative (XY) intervention effectiveness, relative to the power of the original procedure; and (2) although incorporating *both* AB and XY randomization components into the design (as in the present three randomized factor design example) provides a double internal-validity enhancement, the resulting power is exactly the same as that associated with incorporating only one of these additional randomization components (i.e., either AB or XY).

Investigation 5: Randomized Intervention Order for the Single-Case Crossover AB Design

The crossover design is a standard investigative strategy in conventional-group educational intervention research (see, for example, Jones & Hall, 1982; and Levin et al., 1990, Exp. 1). With a crossover design it is possible to compare two intervention conditions (or an intervention and a nonintervention control condition) in two independent groups that also receive both intervention conditions in counterbalanced orders. Although various single-case designs (e.g., the alternating treatment design) allow for each case to receive two or more interventions, the within-case structuring and/or rapid alternation of treatments does not provide an adequate parallel to capture the essence of the crossover design. With a little tweaking, however, the present order-randomization approach can be adapted to capture that essence.

With A and B representing two different interventions, the present order-randomization modification of Marascuilo and Busk's (1988) model has all the apparent trappings of a crossover design. However, adding a straightforward

order-randomization component to that model may not adequately fit a single-case researcher's crossover-design bill. Specifically, randomizing the intervention order independently for all participants (or other units) in the Marascuilo-Busk model does not guarantee that an equal number of participants will receive the two orders, AB and BA—something that is desirable, if not essential, for producing a study that is completely counterbalanced with respect to the order of intervention administration. In fact, in the extreme, a simple randomization scheme could actually result in all participants receiving the same order of intervention administration. In a single-case intervention study with a small number of cases, that situation is not as unlikely as it may initially appear. For example, with $N = 2$ cases it will happen half the time; with $N = 3$ it will happen 25% of the time; and with $N = 4$, it has a 12½% chance of occurring. It should also be recognized that it is not possible to have *complete* (i.e., perfect) order counterbalancing with an odd number of participants.

Consequently, a potentially useful alternative is a crossover design that is completely counterbalanced with respect to the order in which the two different interventions are administered. Implementing such a procedure perfectly controls for potential contaminating effects associated with the two different intervention orders (AB and BA) and therefore eliminates order effects as an internal validity concern. This can be accomplished with a *restricted randomization* scheme, the Type I error and power characteristics of which are explored next in the context of Investigation 5.

Method

In this investigation we examined the effect on Type I error and power characteristics of restricting the dual-randomization scheme to ensure a balance between cases assigned to crossover design orders AB and BA. Specifically, a restricted dual-randomization crossover design (henceforth referred to as *restricted*) with 15 observations and $k = 5$ potential start points for each case randomly selected from observations 6 through 10 was examined for conditions with 2, 3, 4, 5, and 6 cases. For conditions with an even number of participants the number assigned to AB was restricted to equal the number assigned to BA, resulting in an augmented multiplier factor of $\binom{N}{x} = N! / x!(N-x)!$ to the k^N potential intervention start-point randomization outcomes (or $\prod_{i=1}^N k_i$ when the number of potential intervention start points differs across cases), where N is the total number of cases and x is the number of cases that are to be randomly assigned to each of the two administration orders. For an odd number of

IMPROVED RANDOMIZATION TESTS

participants the number assigned to AB was restricted to equal the number assigned to BA, plus or minus 1. In the latter (odd number) case, because of the dual-randomization process of: (1) randomly determining which order, AB or BA, was to be associated with the larger number; and (2) randomly assigning the two orders to participants, this resulted in an augmentation factor of $2\binom{N}{x} = 2\left[N!/x!(N-x)!\right]$ (see Levin et al., 2014, p. 192). Effect sizes were varied from 0 to 3 in increments of .5, again the autocorrelation was set to 0 or .3, and one-tailed $\alpha = .05$ tests were conducted.

Results

Results from the conditions where the autocorrelation is 0 are shown in Figure 11, whereas those for an autocorrelation of .3 are shown in Figure 12. For comparative purposes, results from the unrestricted-dual randomization designs (henceforth referred to as *unrestricted*) of Investigation 3a are also included in those two figures. In Figures 11 and 12 it is clear that for all sample sizes the restricted-randomization tests yielded empirical Type I errors (i.e., when the effect size was 0) that corresponded with their nominal .05 values. Although it is evident from Figures 11 and 12 that the restricted-randomization crossover-design powers are uniformly lower than the corresponding unrestricted-randomization powers, the difference between the two becomes less and less noticeable with increases in sample size. With N s of 5 and 6, for example, the power differences are negligible for all practical purposes. At the same time, it should be pointed out that even at the smaller sample sizes the restricted-randomization crossover-design powers are respectable. To wit, in Investigation 3a it was indicated that with an autocorrelation of .3 and $N = 3$ participants, the unrestricted-randomization test's power for detecting an effect size of $d = 1.5$ was equal to .90 (reproduced in Figure 11); and as may also be seen in Figure 11, for the same set of parameters the restricted-randomization crossover-design test's power is .865.

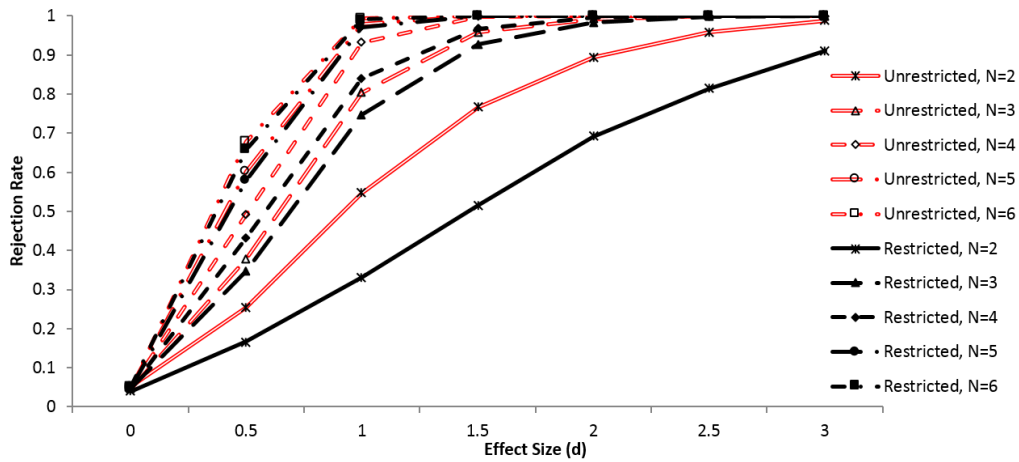


Figure 11. Investigation 5: Comparison ($\alpha = .05$, one-tailed) of randomization tests for the Restricted Dual and Unrestricted Dual AB randomized crossover designs replicated across N cases. The rejection rate of the null hypothesis is shown as a function of effect size and N , for a 15 observations design with 5 potential intervention start points designated from between the 6th and 10th observations inclusive and an autocorrelation of 0.

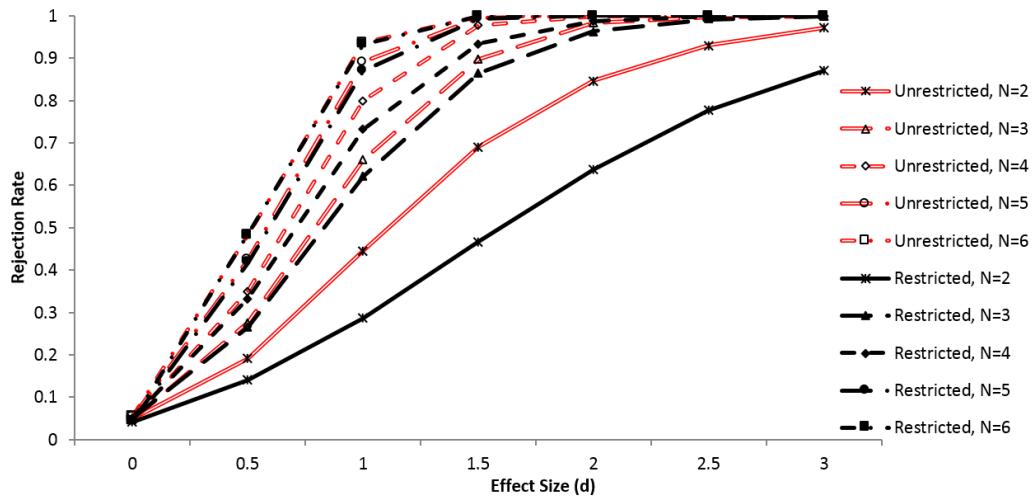


Figure 12. Investigation 5: Comparison ($\alpha = .05$, one-tailed) of randomization tests for the Restricted Dual and Unrestricted Dual AB randomized crossover designs replicated across N cases. The rejection rate of the null hypothesis is shown as a function of effect size and N , for a 15 observations design with 5 potential intervention start points designated from between the 6th and 10th observations inclusive and an autocorrelation of .3.

Investigation 6: Randomized Intervention Order for the Single-Case ABAB Design

In Investigation 5 the simulations were extended to four-phase ABAB designs (also referred to as *reversal* or *operant* designs—see, for example, Kratochwill & Levin, 2010). More specifically, Type I error and power were examined for Onghena's (1992) randomized intervention start-point ABAB design (Single) and a combined randomized intervention start-point plus random-order (ABAB versus BABA) design (Dual), with the dual approach enhancing the ABAB design's internal validity by virtue of its controlling for potentially confounding order effects.

Method

The effect of case replications (more participants) on power was examined for a design with 23 observations and a minimum of 5 observations in each of the four phases, which implies that the number of possible permutations for one case is 20 for the single-randomized design (for computational details, see Onghena, 1992) and 40 for the dual-randomized design. The simulations included 1, 2, 3, or 4 participants, effect sizes that varied from 0 to 3 in increments of .5, and an autocorrelation of 0 or .3. Sample sizes greater than 4 were not investigated because ABAB designs provide more intervention-effect information per case than AB designs and thus they tend to be replicated across fewer participants. Thus, the value in extending the study to larger numbers of participants was judged not to warrant the increased computational time that would have been required. All tests (based on the average of the two B-phase observations minus the average of the two A-phase observations) were conducted with a one-tailed Type I error probability of .05. In that regard, it should be mentioned that the present simulations are based on the weighted (by the number of outcome observations, O) A- and B-phase means [i.e., $(O_{A1}M_{A1} + O_{A2}M_{A2})/(O_{A1} + O_{A2})$ and $(O_{B1}M_{B1} + O_{B2}M_{B2})/(O_{B1} + O_{B2})$] whereas Gafurov and Levin's (2014) *ExPRT* program calculations are based on the unweighted means $[(M_{A1} + M_{A2})/2$ and $(M_{B1} + M_{B2})/2]$. Power differences attributable to the two weighting schemes *per se* should be minimal for the set of parameters that were specified for the present simulations, however.

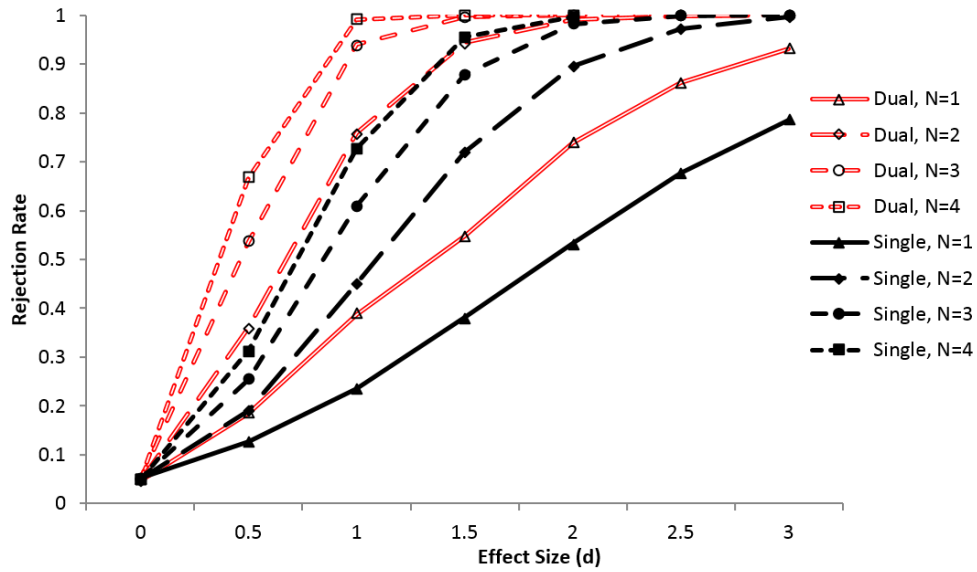


Figure 13

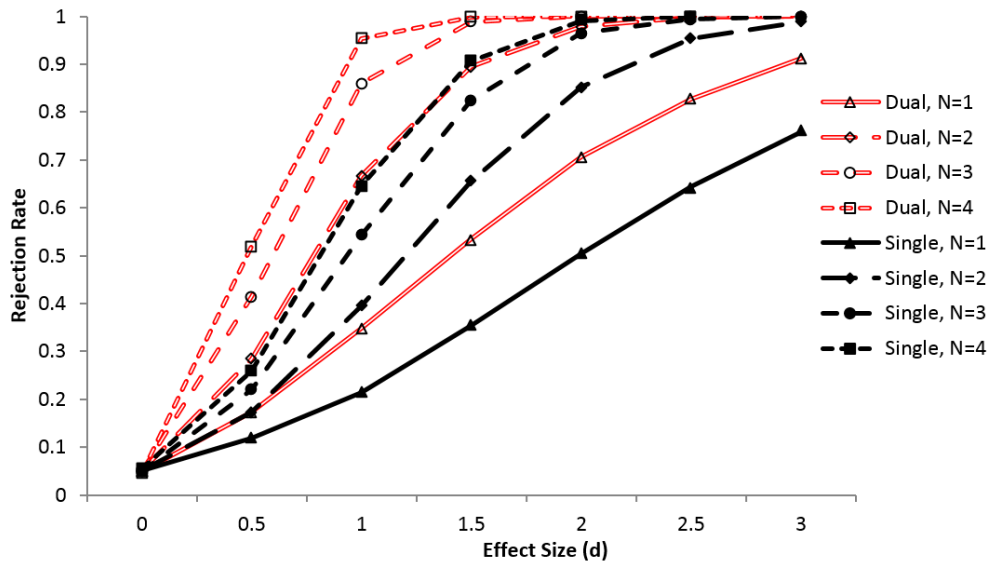


Figure 14

Figures 13 and 14. Investigation 6: Comparison ($\alpha = .05$, one-tailed) of randomization tests for the Single and Dual randomized ABAB designs replicated across N cases. The rejection rate of the null hypothesis is shown as a function of effect size and N , for an autocorrelation of 0 (Figure 13) or .3 (Figure 14), and a 23 observations design with a minimum of 5 observations in each of the four phases. The resulting number of possible randomizations is 20 for the Single randomization scheme and 40 for the Dual randomization scheme.

Results

Results from the conditions where autocorrelation was 0 are shown in [Figure 13](#), while those for an autocorrelation of .3 are shown in [Figure 14](#). As was true for the AB designs, once again the present dual-randomization scheme greatly overpowers the single-randomization scheme. For example, with an autocorrelation of .30 an effect size given by $d = 1.5$, and an $N = 2$ design, single-randomization ABAB power is equal to .66 whereas dual-randomization ABAB power is .895—a nontrivial power difference of almost .24. For the single-randomization scheme to achieve comparable power (.91) to that of the dual-randomization scheme (.895) would require twice as many participants, namely $N = 4$.

Investigations 7 and 8: The Single-Case AB Design Revisited

What follow are two additional AB design investigations, both of which follow directly from colleagues' concerns about data characteristics of the simulations reported thus far. One such concern focuses on the series lengths associated with all of the simulations conducted so far and the other focuses on the distributional characteristics of the outcome measure that comprises all of those simulations. These two concerns are addressed in Investigations 7 and 8, respectively.

Investigation 7

In a recent survey of single-case intervention research reported in 21 journals and based on 809 cases during the year 2008, Shadish and Sullivan (2011) reported that the modal and median series length per case consisted of 20 total observations. The positively skewed distribution had a mean of 27.0 and range of 2 to 160. Approximating from Shadish and Sullivan's frequency histogram ([Figure 2](#)), one can estimate that 23% of the cases had series lengths in the 20-29 range, with 16% in the 30-39 range, 6% in the 40-49 range, and 5% that were 50 or more. Moreover, it is not difficult to locate single-case intervention studies in recent years that included 50 or more outcome observations per case—see, for example, Lucynski, Hanley, & Rodriguez (2014), with 6 children and approximately 50 observations per child; Pellecchia et al. (2011), with 8 children and 60 or more observations per child; Hanley, Jin, Vanselow, & Hanratty (2014), with 3 children and approximately 70 observations per child; and Donaldson,

Trahan, & Kahng (2014), with 1 adult exhibiting dementia and approximately 130 observations.

In the present Investigation 1, the simulation consisted of 30 outcome observations; in Investigation 2, the range spanned from 20 to 150; in Investigations 3 and 4 there were 15 outcome observations; in Investigation 5 there were 15 and 30; and in Investigation 6 there were 23. Therefore, the series lengths for the present simulations do not seem too far out of line with those of single-case intervention studies that are being reported in the literature, where at least half of them include at least 20 observations (Shadish & Sullivan, 2011).

Why, in the first place, was a series as long as 30 decided upon for our Investigations 1 and 5? The answer is simple with respect to the primary focus of the study. Specifically, at least 21 observations (i.e., 20 potential intervention points with at least one baseline observation and one intervention observation) are required to compare Edgington's (1975) single randomization-test procedure and the present dual modification based on a one-tailed α of .05. We settled on 30 total observations to provide at least 5 baseline observations and 5 intervention observations, thereby obtaining some degree of stability in those two series.

That said, in Investigation 7 we examined whether the already reported power difference favoring the dual- over the single-randomization approach would generalize to shorter—in fact, very short—series ($N < 10$), as was analogously examined by Levin et al. (2011) in their short series Investigation 2's AB design.

Method Here, the simulation parameters and procedures of Investigation 3 were again selected and applied to three short-series conditions. Power for each of these conditions was assessed for the single- and dual-randomization test procedures ($\alpha = .05$, one-tailed) for both series based on an autocorrelation of 0 and those based on an autocorrelation of .30.

In one condition two cases were included, with 9 outcome observations per case. The first two observations were always in the first phase, the last two observations were always in the last phase, and the intervention start point was randomly chosen from among the middle five observations in the series. In a second condition three cases were included, with 7 outcome observations per case. The first two observations were always in the first phase, the last two observations were always in the last phase, and the intervention start point was randomly chosen from among the middle three observations in the series. The third condition consisted of five cases, with 8 outcome observations per case. The first three observations were always in the first phase, the last three observations were

IMPROVED RANDOMIZATION TESTS

always in the last phase, and the intervention start point was randomly chosen from among the middle two observations in the series.

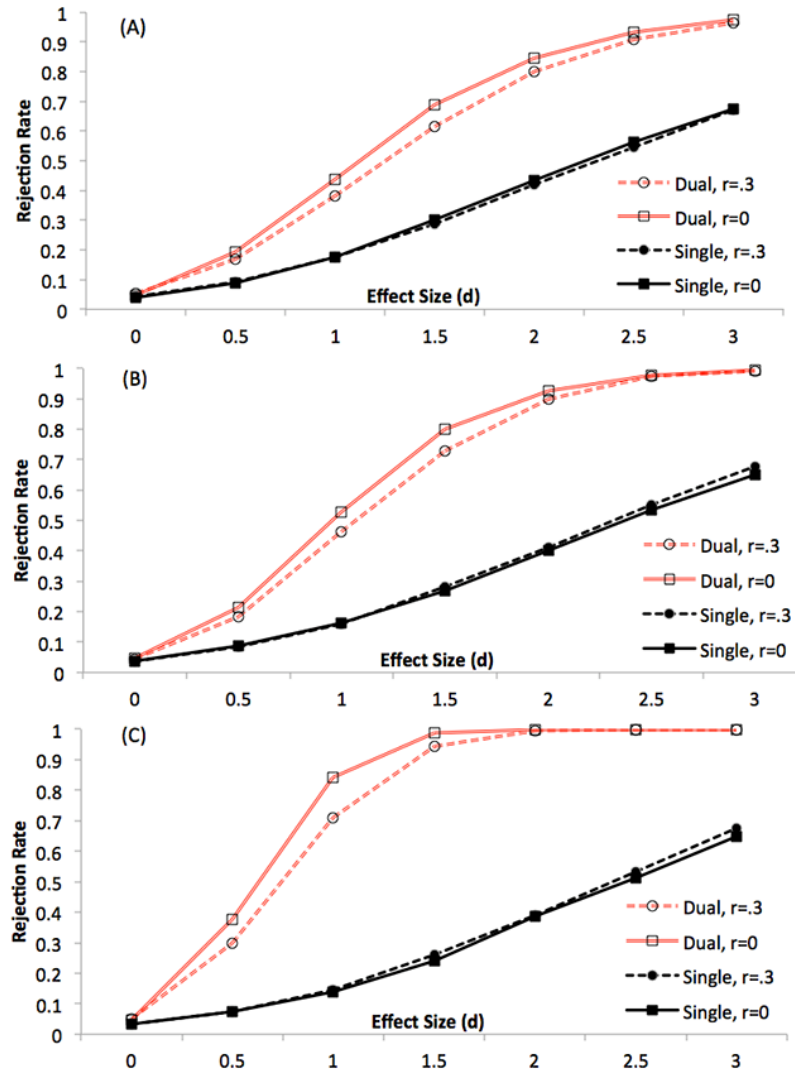


Figure 15. Investigation 7: Comparison ($\alpha = .05$, one-tailed) of randomization tests for the basic AB randomized intervention start-point design (Single) and the randomized intervention start-point plus randomized intervention-order design (Dual). The rejection rate of the null hypothesis is shown as a function of the effect size and level of autocorrelation for: (A) a two-participant design with nine observations each where the start point is randomly assigned to one of the middle five observations, (B) a three-participant design with seven observations each where the intervention start point is randomly assigned to one of the middle three observations, and (C) a five-participant design with eight observations each where the start point is randomly assigned to one of the middle two observations.

Results The results are summarized in the three panels of Figure 15, where it may clearly be seen that, as in Investigation 3a, with the Type I error well controlled, in all three conditions the dual- randomization test's powers by far surpass those of the single-randomization test. A direct comparison of selected dual-over-single power advantages in the long-series Investigation 3a (Figure 4) and the present short-series investigations (Figure 15) is summarized in Table 3, where it should be noted that the advantages in the short-series investigations are comparable to (or larger than) those of the long-series investigations. On that basis, it can be concluded that the appeal of the dual-randomization approach is not restricted to long-series intervention studies. The approach applies equally well, if not better, to intervention studies consisting of a total of 7, 8 or 9 outcome observations.

Table 3. Selected single- versus dual-randomization power comparisons of the present longer (Investigation 3a, Figure 4) and shorter (Investigation 7, Figure 15) series simulations (SL = Series Length, PISP = Number of Potential Intervention Start Points)

<i>N</i>	<i>d</i>	<i>r</i>	Size (SL/PISP)	Single	Dual	Difference
2	2	0.3	Longer (15/5)	0.44	0.85	0.41
			Shorter (9/5)	0.42	0.8	0.38
3	1.5	0.3	Longer (15/5)	0.49	0.9	0.41
			Shorter (7/3)	0.28	0.73	0.45
5	1	0.3	Longer (15/5)	0.45	0.89	0.44
			Shorter (8/2)	0.15	0.71	0.56

As may also be seen in Figure 15, in contrast to the long-series results presented in Figures 3 and 4, throughout the present study, and in previous investigations, the powers associated with the single-randomization approach do not decrease as the autocorrelation increases from 0 to .30. In fact, a slight power increase may be observed for the larger effect sizes in Panels B and C. This same positive relationship between autocorrelation and power for the single-randomization approach was also discovered and noted by Levin et al. (2011) in their short-series Investigation 2. Those authors offered a speculative interpretation of that finding, but a experimental examination of that interpretation remains to be conducted.

Investigation 8

In all of the present simulations, the data were generated assuming that the outcome measure was continuous and normally distributed, whereas in many single-case intervention studies the outcome measures consist of discrete counts or rates. Therefore, to assess whether the power differences favoring the dual-over-single randomization approach would be observed even in an extremely non-normal distribution situation, Investigation 1 was replicated with the only change being that the outcome measure was simulated to be a binary variable as opposed to a continuous one.

Method More specifically, the same algorithms were used to generate the data, but the resulting values were dichotomized such that all values over 1 were recoded as 1 and all values under 1 were recoded as 0. Thus, for conditions without autocorrelation, the baseline observations had a probability of .34 of being a 1 (and .66 of being a 0), whereas the probability of obtaining a 1 in the intervention phase depended on d (e.g., when d equaled 0, 1, 2, 3, 4, and 5, the probabilities of obtaining a 1 were .34, .50, .84, .98, .999, .999997).

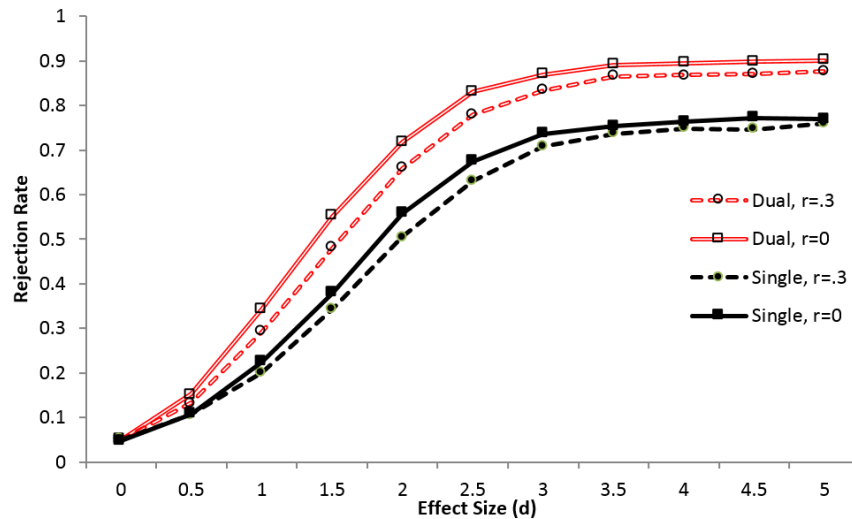


Figure 16. Investigation 8: Comparison ($\alpha = .05$, one-tailed) of randomization tests for the basic AB randomized intervention start-point design (Single) and the randomized intervention start-point plus randomized intervention-order design (Dual), where the outcome is binary and the intervention start point is randomly selected between the 6th through the 25th observations inclusive in a 30-observations study. The rejection rate of the null hypothesis is shown as a function of the effect size and level of autocorrelation.

Results The results of this simulation may be seen in [Figure 16](#). Similar to when the continuous outcome was examined (Investigation 1) the dual-randomization approach consistently leads to greater power than the single-randomization approach, but as would be anticipated, the dichotomization of the outcome lessens the power for each. Also of note, the power estimates reach a ceiling below 1.0, which can be explained by the baseline observations being set so there was a .34 probability of observing the desired behavior. If the baseline probability had been set lower, say to .01, the difference in probabilities between phases could be larger, leading to higher observed maximum powers.

General Discussion

In the eight Monte Carlo investigations reported here, we discovered that in situations where researchers are able to randomize the order in which the phases of single-case AB and ABAB designs (or the interventions themselves in paired-cases designs) can be administered by, for example, simple coin flips, it is clearly advantageous to do so. Order randomization represents a valuable addition to Edgington's (1975) and Onghena's (1992) randomized start-point models, in that it: (1) enhances those designs' internal validity (a methodological improvement); and (2) effectively controls the associated randomization test's Type I error probability, while affording increases in the test's power (a statistical improvement). In many of the instances examined, these power increases were dramatic with respect to a single-case researcher's economic savings. For instance, in Investigation 2's $N = 1$ simulations we found that an AB design with the present dual-randomization scheme could require less than half as many outcome observations as Edgington's original single-randomization scheme. Specifically, as may be seen in [Figure 2](#), for $\alpha = .05$ (one-tailed), an effect size of 2.0, and a series autocorrelation of .3, the dual-randomization approach based on 30 outcomes yields power of .67. In contrast, to achieve similar power with the single-randomization approach requires between 80 and 90 outcome observations. In alternative economic terms, in Investigations 3 and 5 we found that in $N > 1$ investigations, about twice as many participants are required for the single-randomization approach to achieve power equivalent to that of the dual-randomization approach (see [Figures 3-8](#)). Similar dual-over-single randomization power advantages were achieved in the Investigations 4 and 6 randomized pair-members AB design and four-phase ABAB design, respectively. Importantly to single-case researchers from both practical and versatility perspectives, such power advantages were also observed in: (a) short-series designs consisting of as

IMPROVED RANDOMIZATION TESTS

few as seven observations (Investigation 7); and (b) single-case intervention contexts associated with binary, rather than normally distributed, outcome measures (Investigation 8).

Additional Considerations for the Single-Case Crossover Design

Restricted or unrestricted randomization: Which is better? To guarantee order balance (and, therefore, greater internal validity) in single-case AB crossover designs, a restricted dual-randomization scheme must be employed, rather than an unrestricted one. Although the restricted-randomization approach results in powers that are uniformly lower than those associated with an unrestricted-randomization approach, as sample sizes increase beyond $N = 2$ or 3 cases the respective powers of the two designs are quite comparable. So, whenever a researcher is considering the tradeoff between a guaranteed crossover-design balance of intervention administration order (thereby controlling perfectly for order effects), on the one hand, and some degree of increased statistical power, on the other, then: (1) if the former is considered to be relatively more important, the researcher should select the restricted-randomization procedures of Investigation 5; and (2) if the latter wins out as being relatively more important, the researcher should choose the unrestricted-randomization procedures of Investigation 3, especially when the sample size is relatively small (i.e., $N < 3$ or 4 cases).

Controlling for potential confounding factors In actual intervention research studies based on within-subjects designs, in general, and single-case AB crossover designs, in particular (as represented by current Investigation 5), more potentially confounding variables than simple order effects must be taken into account and controlled. That is, between-phase outcome changes may also be the result of other extraneous factors, including: external effects, such as those attributable to history; effects associated with the experimenter or instructor; and effects associated with the participant, such as novelty, Hawthorne, and “John Henry” effects (see, for example, Shadish, Cook, & Campbell, 2002). Such confounding variables can severely compromise an intervention study’s internal validity—namely, that the manipulated intervention *per se* was responsible for between-phase outcome changes—as well as its construct validity. In research now in progress, we are comparing the effects of extraneous factors on internal and statistical-conclusion validity in the present unrestricted and restricted crossover designs.

A random-assignment caveat A few words of operational caution connected to the restricted design crossover design of Investigation 5 should be offered to interventionists who elect to implement that design in their research. Specifically, some researchers are likely to make a critical random-assignment mistake when it comes to implementing the randomization process correctly. With an even number of cases, there should be no problem, in that the researcher would randomly select half of the cases to receive an AB order of intervention administration, with the remaining half receiving the BA order. With an odd number of cases, however, the researcher needs to consider possible assignments where either the AB order or the BA order receives the larger number of cases. To do so, the researcher could go through a two-step randomization process, as follows. In Step 1, the researcher would randomly determine whether the larger number of cases is to receive the AB order or the BA order (e.g., 4 cases if $N = 7$). Then in Step 2, the researcher would proceed as in the previous “even N ” situation, namely randomly selecting the N_1 cases that will be receiving the AB order, with the remaining N_2 cases receiving the BA order. Without the researcher conducting the restricted-randomization procedure in this two-step fashion (or through an analogous completely random-assignment process), subjectivity would enter into the researcher’s decision about which order (AB or BA) receives the one more (or one fewer) case, resulting in the randomization distribution and its associated statistical test being invalid.

Levin and Wampold’s (1999) Simultaneous Pairs Intervention Start-Point Model Revisited

In the present Investigation 4, we examined Levin and Wampold’s (1999) simultaneous pairs, comparative intervention effectiveness hypothesis, with a randomized XY intervention variable included in the randomization-test analysis. In that situation, we found the statistical power of the procedure to be greatly enhanced relative to that of the original Levin-Wampold procedure, for which the randomized intervention factor is not taken into account. We now consider a variation and an extension in conjunction with the present modified procedure.

For the variation, suppose that the A and B phases represent two competing interventions and, as in Investigations 1-3, it is possible to randomize the order in which the two phases are administered (A followed by B or B followed by A). Within each participant pair, it is randomly determined which pair member is assigned the AB administration order and which the BA order (say, $X = AB$ and $Y = BA$). The data are collected and, as in Investigation 4, the comparative

IMPROVED RANDOMIZATION TESTS

intervention effectiveness hypothesis is tested (with the inclusion of the 2^N multiplier associated with the randomized XY factor) on the difference in differences, $(X_{A1} - Y_{B1}) - (X_{B2} - Y_{A2}) = (X_{A1} + Y_{A2}) - (X_{B2} + Y_{B1})$. Note that in this context the interaction actually represents a main effect comparison of Intervention A vs. Intervention B, just as it does in a conventional crossover design. Accordingly, this paired-cases design then becomes conceptually equivalent to the just discussed restricted-order crossover design of Investigation 5, but because of the pairs structure here, for which it is guaranteed that: (1) there will be equal numbers of participants receiving each intervention order; and (2) within each pair, the crossover will occur at exactly the same point in time.

For the extension of the modified Levin-Wampold (1999) simultaneous pairs comparative intervention effectiveness test, now suppose that two equivalently scaled (or commensurable) outcome measures, M_1 and M_2 , are constructed to be differentially sensitive to an intervention; or alternatively, that M_1 is expected to be more responsive to Intervention X than to Intervention Y and M_2 is expected to be more responsive to Intervention Y than to Intervention X—as with Levin’s (1989) experimental illustrations of Campbell and Fiske’s (1959) discriminant validity and Morris, Bransford, and Franks’ (1977) transfer-appropriate processing. The modified dual-randomization procedure to test Levin and Wampold’s comparative intervention effectiveness hypothesis can be readily extended to accommodate the differential outcome-measure effects addition. Specifically, with X and Y representing randomly assigned interventions within each pair, A and B representing baseline and intervention phases (as in Investigation 4), and M_1 and M_2 representing commensurable measures or tests, the data to be analyzed are simply the intervention-by-phase difference-in-differences effect associated with M_1 minus the same effect associated with M_2 , and which amounts to the three-way interaction of intervention by phase by outcome measure. This translates into an assessment of whether whatever differential change from Phase A to Phase B that is produced by the two interventions is the same on the two outcome measures. As with the Investigation 4 test of the two-way intervention-by-phase interaction (i.e., the comparative intervention effectiveness test), the statistical power to test this extended difference would also benefit from the $2N$ multiplier resulting from within-pair randomization of the intervention factor.

Extensions to Other Single-Case Intervention Designs and Situations

Other single-case designs

Research by the present authors is currently in progress to extend the present randomized-order design-and-analysis procedure (combined with randomized intervention start points) to single-case intervention designs other than the AB-type and ABAB designs that were investigated here. Our initial efforts have been targeted at alternating treatment designs (Levin et al., 2012) and multiple-baseline designs. In the former, independently randomizing the alternating A and B intervention phases both within and across participants has been recommended as an internal-validity enhancer (e.g., Kratochwill & Levin, 2010) and incorporating both randomized intervention start points and randomization statistical tests into the process is relatively straightforward. In the latter, although multiple-baseline designs typically include a set of staggered baseline (A) and intervention (B) phases across participants, the present randomized-order approach could be adopted for situations in which, as was discussed here, an initial mandatory A' series of baseline (warm-up or adaptation) observations is included. The approach might also be possible in situations where A represents a standard or basic instructional/behavioral practice and B represents a competing alternative practice.

Other outcome measures

As well as testing for between-phase mean (level) changes, the present randomized-order procedure is similarly applicable to testing for changes in slope (trend) and variance (variability). All such tests are available in Gafurov and Levin's (2014) Excel®-based randomization-test software, which is freely accessible from the Google Drive *ExPRT* (Version 1.2) website, <https://code.google.com/p/expert/>. At the same time, simulation research now in progress (Levin et al., 2014) is assessing the Type I error probabilities and statistical powers of the present combined randomized intervention start-point and randomized-order approaches relative to Koehler and Levin's (1998) randomized intervention start-point approach alone, with respect to tests of slope and variance, in various single-case intervention designs.

Other intervention effect types

It is important to note that in the present eight-investigation set of Monte Carlo simulations, all intervention effects were modeled to represent immediate abrupt changes in the participant's mean level: that is, a constant increase in the participant's series of observations that is coincident with the initial potential intervention point specified by the researcher—or, in the case of the four-phase ABAB design, coincident with the initial potential phase-change (transition) point that was specified for each of the

IMPROVED RANDOMIZATION TESTS

three phase changes. In some of our research in progress we are modeling other types of intervention effects as well, such as immediate gradual effects, delayed abrupt effects, and delayed gradual effects (see, for example, Lall & Levin, 2004). In each of these ongoing simulation studies our goal is to determine whether the present randomized-order approach and associated randomization test afford power benefits that are as impressive in other single-case design contexts (and for other outcome measures) as were discovered in the present AB and ABAB design tests of between-phase changes in level.

Final Comments

Although randomization schemes of the type advocated here may be opposed by single-case intervention researchers who have been steeped in the response-guided tradition (see, for example, Ferron & Levin, 2014), we hope that such schemes will be received more positively by at least some traditional single-case interventionists. In fact, for years many alternating-treatment design users have been diligent in assigning interventions to phases or sessions using a block-randomization process (Kratonchwill & Levin, 2010; for a research example, see Holden, Bearison, Rode, Kapiloff, Rosenberg, & Rosenzweig, 2002). As a cause for further optimism, an increasing number of single-case investigations that have incorporated various forms of randomization design and analysis are appearing in both student dissertations and the published literature (e.g., Ainsworth, 2014; Bardon, Dona, & Symons, 2008; Bice-Urbach, 2015; Bonnet, 2012; Lojkovic, 2014; Regan, Mastropieri, & Scruggs, 2005).

Acknowledgments

We are grateful to Anya Evmenova for providing us with useful research references. Correspondence concerning this article should be addressed to Joel R. Levin at jrlevin@u.arizona.edu.

References

Ainsworth, M. K. (2014). *Effectiveness of the ALL Curriculum to teach basic literacy skills to groups of students with severe disabilities and complex communication needs*. Unpublished doctoral dissertation, George Mason University, Fairfax, VA.

Auerbach, C., & Zeitlin, W. (2014). *SSD for R: An R package for analyzing single-subject data*. New York: Oxford University Press.

Bardon, L. A., Dona, D. P., & Symons, F. J. (2008). Extending classwide social skills interventions to at-risk minority students: A preliminary application of randomization tests combined with single-subject methodology. *Behavioral Disorders, 33*, 141-152.

Bice-Urbach, B. (2015). *Teleconsultation: The use of technology to improve evidence-based practices in rural communities*. Unpublished doctoral dissertation, University of Wisconsin-Madison, Madison, WI.

Billette, V., Guay, S., & Marchand, A. (2008). Posttraumatic stress disorder and social support in female victims of sexual assault: The impact of spousal involvement on the efficacy of cognitive-behavioral therapy. *Behavior Modification, 32*, 876-896.

Bonnet, L. K. (2012). *The effects of point-of-view video modeling on symbolic play actions and play-associated language utterances in preschoolers with autism*. Unpublished doctoral dissertation, George Mason University, Fairfax, VA.

Busk, P. L., & Serlin, R. C. (2005). Meta-analysis for single-case research. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education* (pp. 187-212). Hillsdale, NJ: Erlbaum.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research*. Chicago, IL: Rand McNally.

Donaldson, J. M., Trahan, M. A., & Kahng, S. W. (2014). An evaluation of procedures to increase cooperation related to hoarding in an older adult with dementia. *Journal of Applied Behavior Analysis, 47*, 410-414.

Edgington, E. S. (1975). Randomization tests for one-subject operant experiments. *Journal of Psychology, 90*, 57-68.

Edgington, E. S. (1980). Validity of randomization tests for one-subject experiments. *Journal of Educational Statistics, 5*, 235-251.

Edgington, E. S. & Onghena, P. (2007). *Randomization tests (4th ed.)*. Boca Raton, FL: Chapman & Hall.

IMPROVED RANDOMIZATION TESTS

Ferron, J., & Jones, P. K. (2006). Tests for the visual analysis of response-guided multiple-baseline data. *Journal of Experimental Education*, 75, 66-81.

Ferron, J. M., & Levin, J. R. (2014). Single-case permutation and randomization statistical tests: Present status, promising new developments. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case intervention research: Methodological and statistical advances* (pp. 153-183). Washington, DC: American Psychological Association.

Ferron, J. M., Moeyaert, M., Van den Noortgate, W., & Beretvas, S. N. (2014, June). Estimating causal effects from multiple-baseline studies: Implications for design and analysis. *Psychological Methods*. Advance online publication. <http://dx.doi.org/10.1037/a0037038>.

Ferron, J., & Sentovich, C. (2002). Statistical power of randomization tests used with multiple-baseline designs. *Journal of Experimental Education*, 70, 165-178.

Ferron, J. & Ware, W. (1994). Using randomization tests with responsive single-case designs. *Behavior Research and Therapy*, 32, 787-791.

Ferron, J., & Ware, W. (1995). Analyzing single-case data: The power of randomization tests. *Journal of Experimental Education*, 63, 167-178.

Fisch, G. S. (2001). Evaluating data from behavioral analysis: Visual inspection or statistical models? *Behavioural Processes*, 54, 137-154.

Gafurov, B. S., & Levin, J.R. (2014, Apr.). *ExPRT (Excel7 Package of Randomization Tests): Statistical Analyses of Single-Case Intervention Data* (Version 1.2). Downloadable from <https://code.google.com/p/expirt/>.

Hanley, G. P., Jin, C. S., Vanselow, N. R., & Hanratty, L. A. (2014). Producing meaningful improvements in problem behavior of children with autism via synthesized analyses and treatments. *Journal of Applied Behavior Analysis*, 47, 16-36.

Heyvaert, M., & Onghena, P. (2014). Randomization tests for single-case experiments: State of the art, state of the science, and state of the application. *Journal of Contextual Behavioral Science*, 3, 51-64.

Holden, G., Bearison, D. J., Rode, D. C., Kapiloff, M. F., Rosenberg, G., & Rosenzweig, J. (2002). The impact of a computer network on pediatric pain and anxiety. *Social Work and Health Care*, 36, 21-33.

Horner, R. H., & Odom, S.L. (2014). Constructing single-case research Designs: Logic and options. In T. R. Kratochwill and J. R. Levin (Eds.), *Single-*

case intervention research: Methodological and statistical advances (pp. 27-51). Washington, DC: American Psychological Association.

Jones, B. F., & Hall, J. W. (1982). School applications of the mnemonic keyword method as a study strategy by eighth graders. *Journal of Educational Psychology*, 74, 230–237.

Kazdin, A. E. (1980). Obstacles in using randomization tests in single-case experimentation. *Journal of Educational Statistics*, 5, 253-260.

Koehler, M. J., & Levin, J. R. (1998). Regulated randomization: A potentially sharper analytical tool for the multiple-baseline design. *Psychological Methods*, 3, 206–217.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). Single-case designs technical documentation. In *What Works Clearinghouse: Procedures and standards handbook (Version 2.0)*. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education*, 34, 26-38.

Kratochwill, T. R., & Levin, J. R. (2010). Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychological Methods*, 15, 122-144.

Kratochwill, T. R., & Levin, J. R. (Eds.). (2014). *Single-case intervention research: Methodological and statistical advances*. Washington, DC: American Psychological Association.

Kratochwill, T. R., Levin, J. R., Horner, R. H., & Swoboda, C. M. (2014). Visual analysis of single-case intervention research: Conceptual and methodological considerations. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case intervention research: Methodological and statistical advances* (pp. 91-125). Washington, DC: American Psychological Association.

Lall, V. F., & Levin, J. R. (2004). An empirical investigation of the statistical properties of generalized single-case randomization tests. *Journal of School Psychology*, 42, 61–86.

Levin, J. R. (1989). A transfer-appropriate-processing perspective of pictures in prose. In H. Mandl & J. R. Levin (Eds.), *Knowledge acquisition from text and pictures* (pp. 83-100). Amsterdam: North-Holland.

IMPROVED RANDOMIZATION TESTS

Levin, J. R. (1994). Crafting educational intervention research that's both credible and creditable. *Educational Psychology Review*, 6, 231-243.

Levin, J. R. (1997). Overcoming feelings of powerlessness in aging researchers: A primer on statistical power in analysis of variance designs. *Psychology and Aging*, 12, 84-106.

Levin, J. R., Evmenova, A. S., & Gafurov, B. S. (2014). The single-case data-analysis *ExPRT (Excel Package of Randomization Tests)*. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case intervention research: Methodological and statistical advances* (pp. 185-219). Washington, DC: American Psychological Association.

Levin, J. R., Ferron, J. M., & Kratochwill, T. R. (2012). Nonparametric statistical tests for single-case systematic and randomized ABAB...AB and alternating treatment intervention designs: New developments, new directions. *Journal of School Psychology*, 50, 599-624.

Levin, J. R., Lall, V. F., & Kratochwill, T. R. (2011). Extensions of a versatile randomization test for assessing single-case intervention effects. *Journal of School Psychology*, 49, 55-79.

Levin, J. R., Levin, M. E., Cotton, J. W., Bartholomew, S., Hasty, K., Hughes, C., & Townsend, E. A. (1990). What do college students learn from and about an innovative vocabulary-learning strategy? In S. A. Biggs (Ed.), *Innovative learning strategies, 1989-1990* (pp. 186-206). Pittsburgh, PA: College Reading Improvement Special Interest Group of the International Reading Association.

Levin, J. R., Marascuilo, L. A., & Hubert, L. J. (1978). *N* = nonparametric randomization tests. In T. R. Kratochwill (Ed.), *Single subject research: Strategies for evaluating change* (pp. 167-196). New York: Academic Press.

Levin, J. R., & Wampold, B. E. (1999). Generalized single-case randomization tests: Flexible analyses for a variety of situations. *School Psychology Quarterly*, 14, 59-93.

Lojkovic, D. (2014, Jan.). *Development and use of a modified texting app to increase instances of independent expressive communication for individuals with moderate to severe intellectual and developmental disabilities*. Paper presented at the annual meeting of the Division on Autism and Intellectual Disabilities (DADD), Clearwater, FL.

Luczynski, K. C., Hanley, G. P., & Rodriguez, N. M. (2014). An evaluation of the generalization and maintenance of functional communication and self-

control skills with preschoolers. *Journal of Applied Behavior Analysis*, 47, 246-263.

Maggin, D. M., Swaminathan, H., Rogers, H. J., O’Keeffe, B. V., Sugai, G., & Horner, R. H. (2011). A generalized least squares regression approach for computing effect sizes in single-case research: Application examples. *Journal of School Psychology*, 49, 301-321.

Manolov, R., Evans, J., Gast, D., & Perdices, M. (Eds.). (2014). Single-case experimental design methodology. Special Issue of *Neuropsychological Rehabilitation*, 24.

Manolov, R., & Solanas, A. (2013). A comparison of mean phase difference and generalized least squares for analyzing single-case data. *Journal of School Psychology*, 51, 201-215.

Marascuilo, L. A., & Busk, P. L. (1988). Combining statistics for multiple-baseline AB and replicated ABAB designs across subjects. *Behavioral Assessment*, 10, 1–28.

Marascuilo, L. A., & Levin, J. R. (1970). Appropriate post hoc comparisons for interaction and nested hypotheses in analysis of variance designs: The elimination of Type IV errors. *American Educational Research Journal*, 7, 397-421.

McAllister, L. W., Stachowiak, J.G., Baer, D. M., & Conderman, L. (1969). The application of operant conditioning techniques in a secondary school classroom. *Journal of Applied Behavior Analysis*, 2, 277-285.

McCleary, R., & Welsh, W. N. (1992). Philosophical and statistical foundations of time-series experiments. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis: New developments for psychology and education* (pp. 41-91). Hillsdale, NJ: Erlbaum.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, 16, 519-533.

Moeyaert, M., Ferron, J., Beretvas, S., & Van den Noortgate, W. (2014). From a single-level analysis to a multilevel analysis of single-case experimental designs. *Journal of School Psychology*, 52, 191-211.

Onghena, P. (1992). Randomization tests for extensions and variations of ABAB single-case experimental designs: A rejoinder. *Behavioral Assessment*, 14, 153–171.

IMPROVED RANDOMIZATION TESTS

Onghena, P., & Edgington, E. S. (1994). Randomization tests for restricted alternating treatments designs. *Behaviour Research and Therapy*, 32, 783-786.

Parker, R. I., & Vannest, K. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy*, 40, 357-367.

Pellecchia, M., Connell, J. E., Eisenhart, D., Kane, M., Schoener, C., Turkel, K., Riley, M., & Mandell, D. S. (2011). We're all in this together now: Group performance feedback to increase classroom team data collection. *Journal of School Psychology*, 49, 411-431.

Regan, K. S., Mastropieri, M. A., & Scruggs, T. E. (2005). Promoting expressive writing among students with emotional and behavioral disturbance via dialogue journals. *Behavioral Disorders*, 33-50.

SAS (2013). *SAS/IML® 13.1 User's Guide*. Cary, NC: SAS Institute Inc.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.

Shadish, W. R., Hedges, L. V., Pustejovsky, J. E., Boyajian, J. G., Sullivan, K. J., Andrade, A., & Barrientos, J. L. (2014). A *d*-statistic for single-case designs that is equivalent to the usual between-groups *d*-statistic. *Neuropsychological Rehabilitation*, 24, 529-553.

Shadish, W. R., Kyse, E. N., & Rindskopf, D. M. (2013). Analyzing data from single-case designs using multilevel models: New applications and some agenda items for future research. *Psychological Methods*, 18, 385-405. DOI: [10.1037/a0032964](https://doi.org/10.1037/a0032964)

Shadish, W. R., & Sullivan, K. J. (2011). Characteristics of single-case designs used to assess intervention effects in 2008. *Behavior Research Methods*, 43, 971-980.

Stocks, J. T., & Williams, M. (1995). Evaluation of single subject data using statistical hypothesis tests versus visual inspection of charts with and without celeration lines. *Journal of Social Research*, 20, 105-126.

Woodfield, T. J. (1988). Simulating stationary Gaussian ARMA time series. *Computer Science and Statistics: Proceedings of the 20th Symposium on the Interface*, pp. 612 - 617.

Regular Articles: **Some General Guidelines for Choosing Missing Data Handling Methods in Educational Research**

Jehanzeb R. Cheema

University of Illinois at Urbana-Champaign
Champaign, IL

The effect of a number of factors, such as the choice of analytical method, the handling method for missing data, sample size, and proportion of missing data, were examined to evaluate the effect of missing data treatment on accuracy of estimation. A methodological approach involving simulated data was adopted. One outcome of the statistical analyses undertaken in this study is the formulation of easy-to-implement guidelines for educational researchers that allows one to choose one of the following factors when all others are given: sample size, proportion of missing data in the sample, method of analysis, and missing data handling method.

Keywords: Missing data, imputation, simulation, listwise deletion, missing value analysis

Introduction

Missing data is an issue that most researchers in education encounter on a routine basis. In survey research there can be many reasons for missing data such as respondents ignoring a few or all questions, questions being irrelevant to the respondent's situation, or inability of survey administrators to locate the respondent. Missing data can also occur in non-survey data, such as experimental and administrative data (Acock, 2005; Brick & Kalton, 1996; Groves et al., 2004). In non-survey samples, missing data can arise due to carelessness in observation, errors made during data entry, data loss due to misplacement etc. Regardless of the reason why data is missing, once it is missing it becomes part of the dataset that is then used by researchers to perform analytical procedures. The quality of such

Dr. Cheema is a Clinical Assistant Professor in the Bureau of Educational Research, College of Education. Email him at jrcheema@illinois.edu.

MISSING DATA GUIDELINES

analytical procedures directly depends on the quality of underlying data which in turn can be affected by the nature of missing data (Allison, 2001; Schafer & Graham, 2002).

Unfortunately there are many different methods of handling missing data which can have profoundly different effects on estimation. For this reason it is important to select the correct missing data handling method that is suited to a researcher's particular circumstances. These circumstances can be expressed as factors, such as sample size, proportion of missing data, method of analysis etc., some of which may fall under the control of the researcher in a given scenario and thus can be manipulated, while others are more difficult to control.

For example, a researcher working with secondary data will likely not find it possible to increase the sample size to offset the effect of missing data but may have flexibility regarding the choice of analytical method. On the other hand, a researcher who is gathering her own data and who is relying on a specific method of analysis to answer her research questions may find it easy to increase her sample size in order to lower the proportion of missing cases. As these illustrations suggest, the scenario under which a researcher handles missing data can vary considerably depending on that researcher's circumstances.

There were many investigations and comparisons of the performance of missing data handling methods, both in general (Afifi & Elashoff, 1966; Graham, Hofer, MacKinnon, 1996; Haitovsky, 1968; Peng, Harwell, Liou, & Ehman, 2009; Peugh & Enders, 2004; Wayman, 2003; Young, Weckman, & Holland, 2011) and in context of specific factors such as proportion of missing data (Alosh, 2009; Knol et al., 2010; Rubin, 1987) and sample size (Alosh, 2009; Rubin, 1987). Because the current study is not a review of the literature, any comprehensive attempt to reproduce that discussion is beyond its immediate scope. For detailed technical aspects including mathematically-intensive proofs and theorems, and application of these methods in various fields including education, see Madow, Nisselson and Olkin (1983), Madow and Olkin (1983), Madow, Olkin, and Rubin (1983), Jones (1996), Groves, Dillman, Eltinge, and Little (2002), and Andridge & Little (2010).

Although several researchers have investigated missing data handling methods, their results were based on various combinations of sample size, proportion of missing data, method of analysis, and missing data handling method. None of the past studies has dealt with all of these factors simultaneously using the same dataset in order to control for data-specific characteristics. For this reason, the findings of these earlier studies cannot be used to construct general guidelines for use with new datasets. This study controls for all of these factors simultaneously, and also expands the range of sample size and proportion of missing data in order

to improve the generalizability of its findings. Furthermore, in this study the missing data handling methods are compared for four analytical methods that are frequently employed in educational research: one sample t test, independent samples t test, two-way ANOVA, and linear multiple regression. Results of these comparisons can be used to correct biases in tests of hypotheses reported in past research that employed improper imputation methods, such as mean imputation, that are well-known to produce biased parameter estimates.

Even though the drawbacks of many missing data handling methods are well-known and have been regularly publicized in leading peer-reviewed journals, researchers in social sciences in general and education and psychology in particular have shown a remarkable resilience in sticking to some of the simpler and most error-prone methods such as listwise deletion, pairwise deletion, and mean imputation (Peng et al., 2006; Peugh & Enders, 2004; Roth, 1994; Schafer & Graham, 2002). There are various reasons for avoiding sophisticated missing data handling methods that range from a lack of expertise in quantitative methodology required for a basic understanding of these methods to the inability to practically implement those methods using specialized software programs due to a lack of programming know-how. A correction of this state of affairs requires a study that specifically targets this population of researchers and that can provide general guidelines for selection of the best missing data handling method under a variety of scenarios. Some prior studies such as Roth (1994) have pointed out the absence of an expansive measurement of bias due to missing data and the gain in efficiency that can be achieved by imputing that data in social science literature, especially psychology, a field from which educational research heavily borrows its quantitative methodology. The same study especially stressed development of guidelines that can be used to choose the best missing data handling technique in a variety of circumstances faced by researchers.

The main objective of this study is to provide educational researchers with general guidelines about which missing data handling method performs best under a variety of combinations of sample size, proportion of missing data, and method of analysis. More specifically, these guidelines will allow the researcher to choose one of the following factors when all others are given: sample size, proportion of missing data, method of analysis, and missing data imputation method.

Method

The analytical procedures presented in this study use two sources of data, a simulated dataset and empirical samples. A description of these datasets and analytical procedures follows.

Data Simulation

The primary source of data used for statistical analyses performed in this study was a simulated dataset. The main reason for using simulated data was to ensure that distributional assumptions governing the methods of analysis applied in this study were not violated. The main concern was that violation of underlying model assumptions for each method of analysis under some conditions and not the others can significantly erode uniformity of the basis on which these methods are compared. A reliable way to avoid this problem was to simulate data that satisfied all underlying assumptions for analytical methods of interest and that at the same time had characteristics that made such data suitable for analysis of real-world problems.

In order to mimic data routinely encountered by educational researchers a dataset with 10,000 cases was simulated which included four continuous and one categorical variable. Because groups of variables are usually investigated because they are related to each other, it is important that the simulated data also mimic such relationships. This was achieved by specifying a variance-covariance matrix that was not unlike what a typical educational researcher may encounter during her research.

The four continuous variables, Y , X_1 , X_2 , and X_3 , were generated in such a way as to simulate weak correlation between Y and X_1 ($r = .3$), moderate correlation between Y and X_2 ($r = .5$), and strong correlation between Y and X_3 ($r = .7$), with the three X 's correlated weakly with each other ($r = .2$). This pattern was adopted to avoid the problem of multicollinearity in linear multiple regression models analyzed in this study. It should be noted that the strength of an association is a relative concept. While a coefficient of correlation of .7 may be considered weak in context of a physical experiment, the same might be considered very strong in context of a social study. Cohen (1992), for instance, suggests .1, .3, and .5 as rule of the thumb for small, medium, and strong correlation. Values of the four continuous variables X_1 , X_2 , X_3 , and Y were drawn from a multivariate normal distribution. For ease of interpretation all continuous variables were specified to have a mean of 0 and standard deviation of 1. Dichotomous predictor Z_1 was

constructed using a uniform discrete distribution with values 0 ($n = 4,945$) and 1 ($n = 5,055$).

Because the assignment of these values to Z_1 is random, this mirrors a situation where a significant mean difference in Y does not exist across levels of Z_1 . In order to construct the opposite scenario where mean differences do exist, Z_2 was constructed to have three levels, with mean Y significantly different between these levels. The three levels of Z_2 were labeled 1 ($n = 1,623$), 2 ($n = 6,823$), and 3 ($n = 1,554$) with mean Y being the largest for group 1 and smallest for group 3. It should be noted that even though this means that the pattern of missing data in Y now depends on Z_2 , such dependency rules out only the missing completely at random (MCAR) assumption and not the relatively less stringent missing at random (MAR) assumption and as the missing values of Y are still independent of their own magnitude, the data cannot be considered as not missing at random (NMAR).

Data Analysis Approach for Simulated Data

The simulated dataset ($n = 10,000$) was used to select 10 sub-samples of size 10, 20, 50, 100, 200, 500, 1000, 2000, 5000, and 10000. Each of these sub-samples was then reduced in size by 1%, 2%, 5%, 10%, and 20% in order to simulate datasets containing missing data. The cases were discarded randomly from each complete sample five times separately in order to make sure that there were no dependencies between samples. Each of the five missing data handling methods were applied to all samples containing missing data under four methods of analysis. These methods of analysis are one sample t test, independent samples t test, two-way ANOVA, and multiple regression.

The main considerations behind the choice of these four methods of analysis is their widespread use among educational researchers and the desire not to restrict the findings of this study to a single method of analysis. These methods represent various modeling regimes encountered routinely by researchers in education. For the independent samples t test, the mean difference in Y over levels of Z_1 , the only categorical predictor with two levels, was analyzed. For two-way ANOVA, both categorical predictors, Z_1 and Z_2 were used as factors of Y . And for multiple regression, Y was specified as a function of the three X 's and Z_1 . Five missing data handling methods were selected for missing data analysis. These methods are listwise deletion, mean imputation, regression imputation, maximum likelihood imputation (ML), and multiple imputation. These methods were chosen because of their ready availability and easy implementation in general statistics packages such as SPSS. Application of these five missing data handling methods under various

MISSING DATA GUIDELINES

sample sizes with data missing in various proportions, and for different methods of analysis forms the core of simulated data analysis.

For each of the four methods of analysis, model parameter estimates and associated tests of hypotheses were obtained generated for the 10 complete and 50 partial samples using each of the five missing data handling methods. In other words, a total of $4 \times (10 + 50) \times 5 = 1,200$ models were fitted. These 1,200 models can be categorized into two groups with the first group comprising of 200 models based on samples that contain no missing data and the second group comprising of 1,000 models based on samples that contain missing data. The model significance for these two groups was then compared using the t statistic for models involving one sample t test and independent samples t tests, and the F statistic for two-way ANOVA and multiple regression models.

For example, the F statistic evaluating model significance for two-way ANOVA under multiple imputation of missing data when the sample size is 100 and proportion of missing data is 5% can be directly compared with the corresponding F statistic for the complete sample containing no missing data ($n = 100$). Such a comparison is fair because after imputation the numerator and denominator degrees of freedom are the same for both F values. Thus, since the two samples are identical in all other respects including power, any fluctuation in the observed value of F can be attributed to the deviation of imputed values from their true counterparts. Such an approach allows an objective evaluation of the effect of an imputation method on the statistic used to test for model significance. For instance if the observed F value increases after imputation of missing data, it means that the observed probability of making a Type I error, i.e. rejecting H_0 when H_0 should not be rejected, has decreased.

In order to compare performance of the 1,000 models based on missing data with their complete-data counterparts, a unitless standardized measure of error, the normalized root mean squared error ($RMSE$) was utilized. $RMSE$ is in essence the average distance of observed error from the true value and can be interpreted as the standard deviation of $X_{Observed}$. This measure thus takes into consideration the absolute size of error. However, $RMSE$ calculated in this way has the same unit of measurement as X . By dividing $RMSE$ with the range of X , the unit of measurement can be removed from $RMSE$. The resulting statistic is called the normalized $RMSE$. The advantage of using normalized $RMSE$ over $RMSE$ is that it can be used to compare error across variables that are not based on the same unit of measurement.

Empirical Sample 1

In order to test the real-world applicability of simulated results, a large scale dataset with variables having characteristics similar to those used in the simulated data was utilized. This empirical data was obtained from U.S. portion of the Program for International Student Assessment (PISA) (NCES, 2003) which is an assessment of literacy in mathematics, reading, and science of 15-year old students ($n = ,456$). The questionnaire for this survey was the basis for a large number of variables, some of which are comparable to those simulated in this study. The primary idea behind using an empirical sample was to test the effectiveness of guidelines constructed on the basis of simulated data. The variable selection was based on similarity of characteristics of these variables with their simulated counterparts.

The dependent variable was math achievement which was distributed normally, measured on a continuous scale, and ranged between 200 and 800. Three continuous variables were chosen as predictors of math achievement on the basis of similarity between the variance-covariance matrix of these predictors and that of the simulated continuous variables. These predictors are reading achievement, math anxiety, and the index of home educational resources. Reading achievement was normally distributed and ranged between 200 and 800. Math anxiety is a measure of anxiety felt by a student when engaged in math-related tasks. This variable was measured on a continuum, was normally distributed, and standardized to have a mean of 0 and standard deviation of 1.

Home educational resources measured educational resources owned by a student's household and can be roughly thought of as a component of the student's socioeconomic status. The variable was also standardized to have a mean of 0 and standard deviation of 1. A comparison between the variance-covariance matrices of simulated and empirical predictors showed slight differences in magnitude. However, what is more important to note is the similarity in the pattern of relationship among the four variables which showed that math achievement was correlated somewhat weakly with home educational resources ($r = .3$), moderately with math anxiety ($r = -.4$), and strongly with reading achievement ($r = .8$). This pattern was not very different from that simulated for Y and its three continuous predictors. Similarly, the inter-predictor correlations presented were also weak like their simulated counterparts ranging between $-.3$ and $.3$.

The observed deviation between these two variance-covariance structures emphasizes the practical difficulty associated with obtaining empirical datasets which possess exact distributional characteristics that a researcher may require. In addition to continuous variables a categorical predictor, gender was selected from the PISA 2003 dataset. Gender has two categories: male, $n = 2,740$; and female,

MISSING DATA GUIDELINES

$n = 2,715$. One case had a missing value for gender reducing the maximum number of observations available for analysis from 5,456 to 5,455.

For the purposes of this study, the approach used for simulated dataset was replicated with PISA data. This allows us to compare estimation results with and without missing data imputation. For analysis, we predicted math achievement from its predictors using a linear multiple regression equation.

The empirical variables were used to evaluate the effectiveness of missing data handling guidelines formed with simulated data. A portion of the empirical dataset was designated as missing and was then analyzed using the same missing data handling methods that were employed for simulated data analysis. This involved selecting an appropriate analytical method, estimating model parameters, and then comparing the estimation results for complete dataset with its incomplete and imputed counterparts in order to evaluate whether the differential effects of missing data handling methods.

Empirical Sample 2

A smaller empirical dataset was employed in order to evaluate the effectiveness of missing data handling methods for small datasets. This data comes from the Population and Housing portion of decennial U.S. Census published by the U.S. Census Bureau (2000). The data chosen for this example is for the states of Virginia and Wisconsin and includes the percentage of individuals in each county with at least a four year college degree for the year 2000. The dataset consists of 207 counties (Virginia, $n = 135$; Wisconsin, $n = 72$).

As with empirical sample 1, the objective of using this sample was to illustrate the effect of missing data handling methods on accuracy of estimation. This was accomplished by specifying a portion of the data as missing using a subset of the missing data percentages used for the simulated dataset. Next, missing data were imputed and the parameter estimates obtained with and without imputation were compared in order to evaluate the effect of various missing data handling methods. In contrast to empirical sample 1 for which a relatively advanced method of analysis viz. multiple regression was employed, for empirical sample 2 a simpler method viz. independent samples t test was used to ensure a broader coverage of analytical methods chosen for this study.

Results

Simulated Data

Results of analytical procedures described in the method section for the simulated dataset are presented in this section. In order to see the association between original and imputed data, Pearson coefficient of correlation was calculated between original data and imputed data separately for each imputation method. These correlations were significantly different from zero at 5% level of significance and showed a general decreasing trend in magnitude as the percentage of missing data increased.

Furthermore, the correlations tended to be stronger for maximum likelihood (ML) imputation and multiple imputation methods as compared to mean imputation and regression imputation. When proportion of missing data was 5% or less, almost without exception, all imputation methods produced correlations between original and imputed data that were in excess of .95. Only for sample sizes that were less than 50 with percentage of missing data exceeding 5% did we observe somewhat weaker correlations, in one case falling as low as .74. Mean imputation seemed to work well as long as the percentage of missing data was 10% or less but the correlation between mean imputed and original data fell quickly regardless of sample size as this percentage exceeded 10%. The mean correlation (i.e. correlations averaged over sample size and percentage of missing data) between original and imputed data for mean imputation, regression imputation, maximum likelihood imputation, and multiple imputation were .95, .96, .98, and .98 respectively, suggesting that such correlation was strongest for ML and multiple imputation methods and weakest for mean imputation. However, it should be noted that the difference in magnitude of these correlations is very small.

An examination of normalized *RMSE* values (see Figure 1) showed that multiple imputation was the best missing data handling method because it produced the smallest normalized *RMSE* for all four methods of analysis, one sample *t* test, independent samples *t* test, two-way ANOVA, and multiple regression. For one sample *t* test, all imputation methods performed better than listwise deletion although the difference between listwise deletion and mean imputation was small. For independent samples *t* test, listwise deletion did not perform very well but mean imputation did. Furthermore, for independent samples *t* test, the performance of mean imputation and ML imputation was almost the same. For two-way ANOVA, listwise deletion was as good as ML imputation and better than regression imputation and mean imputation, the latter being the most error-prone method. For

MISSING DATA GUIDELINES

multiple regression, regression imputation worked almost as well as multiple imputation which produced the smallest normalized $RMSE$, listwise deletion and ML imputation behaved similarly, and mean imputation was clearly inferior to all other missing data handling methods.

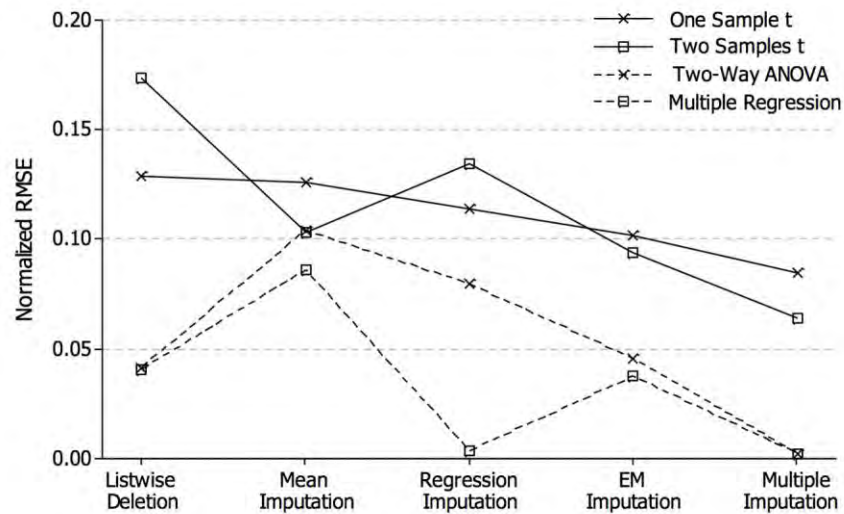


Figure 1. The average effect of missing data handling method on accuracy of estimation for various methods of analysis

The reason why regression imputation performed so well when the analytical method was multiple regression was that using regression-imputed data in a regression equation, when the variables used for imputation and model estimation are the same, is akin to fitting a regression equation twice to predict the same dependent variable. It is important to note here that the results presented in [Figure 1](#) were averaged over sample size and proportion of missing data and therefore cannot be used to evaluate the partial effect of these two factors.

In fact, such averaging contributes to observance of some contradictory results. For example, we see in [Figure 1](#) that mean imputation does not work very well in case of one sample t test but does work well for independent samples t test even though both methods involve a similar kind of dependence on the sample mean of Y and its standard error. For this reason, it is essential that we disaggregate the results in order to clarify the partial effects of sample size and proportion of missing data.

Disaggregated results showed that for one sample t test: with small samples ($n \leq 50$), ML imputation worked best whether the proportion of missing data was low ($m \leq 5\%$) or high ($m > 5\%$); with medium samples ($50 < n < 1,000$), multiple imputation worked best regardless of proportion of missing data; and with large samples ($n \geq 1,000$), ML imputation works best when proportion of missing data was low and multiple imputation worked best when proportion of missing data was high. It should be noted here that even though we have identified the best missing data method under various conditions, in practical terms the increase in efficiency gained due to applications of that best method may be too small to justify such application.

Power comparisons for the four methods of analysis suggested that with listwise deletion and medium effect sizes as defined by Cohen (1992): one sample t test achieved a power of .8 at sample sizes between 20 and 50 for any proportion of missing data ranging between 1% and 20%; independent samples t test achieved a power of .8 at sample sizes between 100 and 200 for any proportion of missing data ranging between 1% and 20%; 2×3 ANOVA achieved a power of .8 at sample sizes between 200 and 500 for any proportion of missing data ranging between 1% and 20%; and multiple linear regression with one set of four predictors achieved a power of .8 at sample sizes between 50 and 100 for any proportion of missing data ranging between 1% and 10% and, at sample sizes between 100 and 200 when the proportion of missing data was 20%. It should be noted that for the four imputation methods, power values at all sample sizes were exactly identical to those of the complete data because after imputation sample sizes are at their maximum.

Statistical results for efficiency gains are summarized as a decision tree in Table 1. Out of the 24 possible situations listed in Table 1 based on various combinations of method of analysis (one sample t test, independent samples t test, two-way ANOVA, multiple regression), sample size (small, medium, large), and proportion of missing data (low, high), relative to listwise deletion, in 15 cases (62.5%) the best method was multiple imputation, in seven cases (29.1%) the best method was maximum likelihood imputation, in only one case (4.2%) the best method was regression imputation, and in only one case (4.2%) the best method was mean imputation. However, the increase in efficiency gained in each of these 24 cases was not the same. For example when multiple regression is the method of analysis, sample size is small, and proportion of missing data is high, the gain in accuracy, defined as the reduction in normalized root mean squared error between the most efficient missing data handling method (multiple imputation in this scenario) and listwise deletion is only about 1%. Thus, in terms of the time and effort required for application of multiple imputation of missing data a researcher

MISSING DATA GUIDELINES

may not find it worthwhile to implement missing data imputation at all rather relying on listwise deletion and be content with the corresponding 1% loss in accuracy that could have been gained otherwise.

Table 1. Summary of gain in estimation accuracy from application of missing data handling methods for various methods of analysis

Sample size ^a	One Sample <i>t</i> Test						Independent Samples <i>t</i> Test					
	<i>Small</i>		<i>Medium</i>		<i>Large</i>		<i>Small</i>		<i>Medium</i>		<i>Large</i>	
Incidence of missing data ^b	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High
Most efficient data handling method ^c	ML	ML	MI	MI	ML	MI	ML	MI	R	MI	MI	MI
Gain in accuracy ^d	0.07	0.01	0.05	0.11	0.01	0.07	0.06	0.01	0.07	0.04	0.15	0.24

Sample size	Two-Way ANOVA						Multiple Regression					
	<i>Small</i>		<i>Medium</i>		<i>Large</i>		<i>Small</i>		<i>Medium</i>		<i>Large</i>	
Incidence of missing data	Low	High	Low	High	Low	High	Low	High	Low	High	Low	High
Most efficient handling method	MI	MI	MI	MI	MI	MI	ML	MI	ML	MI	ML	MI
Gain in accuracy from imputation	0.04	0.03	0.03	0.09	0.02	0.10	0.04	0.01	0.03	0.12	0.02	0.10

Note. EM = Expectation maximization imputation. M = mean imputation. MI = multiple imputation. R = Regression imputation; ^aSmall, $n \leq 50$; Medium, $50 < n < 1,000$; Large, $n \geq 1,000$; ^bLow, missing $m \leq 5\%$; High, missing $m > 5\%$; ^cMost efficient data handling method is the one that produces smallest normalized root mean squared error; ^dGain in accuracy is measured as the reduction in normalized root mean squared error between the most efficient missing data handling method and listwise deletion. When multiplied by 100 this gain can be interpreted as a percentage.

Empirical Sample 1

In order to allow comparison with simulated data results, a multiple regression equation was used to predict math achievement from reading achievement, math anxiety, home educational resources, and gender. Results for the full dataset and the datasets based on various missing data handling methods are presented in Tables 2 and 3 under low, $m = 5\%$ ($n = 5,182$) and high, $m = 10\%$ ($n = 4,910$) missing data conditions respectively.

Table 2. Predicting math achievement: multiple regression results with 5% missing data under various missing data handling methods using PISA 2003 data

Predictors	Partial Slope Coefficient Estimates					
	Full Data	Listwise Deletion	Mean Imputation	Regression Imputation	EM Imputation	Multiple Imputation
Intercept	47.20***	48.12***	72.40***	48.39***	48.12***	48.29***
Gender ^a	30.57***	30.26***	28.31***	30.10***	30.23***	30.18***
Home educational resources	0.68	0.79	0.48	0.76	0.79	0.79
Math anxiety	-11.48***	-11.38***	-11.16***	-11.47***	-11.38***	-11.43***
Reading achievement	0.85***	0.85***	0.80***	0.84***	0.85***	0.84***
Model summary						
<i>F</i>	7676.57***	7229.55***	5494.187***	7640.99***	8056.30***	7669.93***
<i>R</i> ²	.849***	.848***	.801***	.849***	.855***	.849***
<i>Power</i>	1.000	1.000	1.000	1.000	1.000	1.000

Note. *n* = 5,455. *F* = Observed *F* from regression ANOVA. *R*² = proportion of explained variance; ^aReference category is female; * *p* < .05; ** *p* < .01; *** *p* < .001

Table 3. Predicting math achievement: multiple regression results with 10% missing data under various missing data handling methods using PISA 2003 data

Predictors	Partial Slope Coefficient Estimates					
	Full Data	Listwise Deletion	Mean Imputation	Regression Imputation	EM Imputation	Multiple Imputation
Intercept	47.20***	48.36***	88.21***	47.24***	48.36***	48.84***
Gender ^a	30.57***	30.35***	27.53***	30.65***	30.35***	30.22***
Home educational resources	0.68	0.82	0.58	0.89*	0.82	0.86
Math anxiety	-11.48***	-11.72***	-11.04***	-11.73***	-11.72***	-11.83***
Reading achievement	0.85***	0.84***	0.77***	0.85***	0.84***	0.84***
Model summary						
<i>F</i>	7676.57***	6927.02***	4633.834***	7667.110***	8462.10***	7638.26***
<i>R</i> ²	.849***	.850***	.773***	.849***	.861***	.849***
<i>Power</i>	1.000	1.000	1.000	1.000	1.000	1.000

Note. *n* = 5,455. *F* = Observed *F* from regression ANOVA. *R*² = proportion of explained variance; ^aReference category is female; * *p* < .05; ** *p* < .01; *** *p* < .001

These results show that with the exception of mean imputation, all missing data handling methods produce regression parameter estimates and model statistics such as *R*² and overall *F* for regression ANOVA that are very similar to their full data counterparts. Almost without exception, the results of tests of hypothesis from each of the models presented in Tables 2 and 3 are identical. The only exception is when regression imputation is used under the 10% missing data condition and

MISSING DATA GUIDELINES

where home educational resources turns out to be a significant predictor of math achievement ($B = 0.87, p = .048$). This observation of an exception underscores the importance of relying on more than one missing data handling method when percentage of missing data is large (exceeds 5%) as also suggested by Raymond and Roberts (1997).

Although the R^2 values presented in Tables 2 and 3 suggest that regression imputation and multiple imputation methods provide effect size estimates that closely match their full data counterparts, it should be noted that the resulting gains in efficiency are very small compared to listwise deletion ($< 1\%$). In other words, for the large sample ($n = 5,455$) used in this example, listwise deletion is almost as good a choice as the best missing data imputation method. The next step is to see if this result also holds when the sample size is relatively much smaller.

Empirical Sample 2

For the small sample illustration U.S. Census Bureau (2000) data were used. This dataset was used to test for mean difference in percentage of individuals, twenty-five years or older, with college degrees at county level between the states of Virginia and Wisconsin. The sample size was 207 (Virginia, $n = 135$; Wisconsin, $n = 72$). The independent samples t test results based on various missing data handling methods are presented in Tables 4 and 5 under low, $m = 5\%$ ($n = 197$) and high, $m = 10\%$ ($n = 186$) missing data conditions respectively.

These results show that, in terms of effect size, best results are obtained with listwise deletion ($d = .26$) and ML imputation ($d = .26$) when the proportion of missing data is small, and with mean imputation ($d = .25$) when the proportion of missing data is large. Power statistics suggest a small increases in power, from .915 to .926 (gain = 1.2%) when proportion of missing data is small and from .894 to .926 (gain = 3.8%) when proportion of missing data is large. In terms of the effect on test statistic, results were not consistent for all missing data handling methods. For instance, with 5% missing data the null hypothesis of no significant mean difference in percentage of individuals, twenty-five years or older, with college degrees at county level between the states of Virginia and Wisconsin was rejected under listwise deletion ($t = 2.08, p = .039$), mean imputation ($t = 2.19, p = .030$), ML imputation ($t = 2.09, p = .038$), and multiple imputation ($t = 1.87, p = .038$), but not under regression imputation ($t = 1.84, p = .067$). With 10% missing data, this same null hypothesis was rejected under mean imputation ($t = 2.02, p = .044$) and regression imputation ($t = 2.18, p = .031$) but not under listwise deletion ($t = 1.82, p = .071$), ML imputation

Table 4. Independent samples *t* test results for education attainment with 5% missing data under various missing data handling methods using the 2000 census data

	Summary statistics					
	Virginia			Wisconsin		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Full data	135	19.70	11.47	72	17.22	6.16
Listwise deletion	128	19.87	11.38	69	17.26	6.28
Mean imputation	135	19.87	11.08	72	17.26	6.14
Regression imputation	135	19.67	11.17	72	17.42	6.36
EM imputation	135	19.83	11.08	72	17.33	6.15
Multiple imputation	135	19.76	11.35	72	17.17	6.44

	<i>t</i> test statistics						
	<i>t</i>	<i>df</i>	<i>p</i>	ΔM	<i>SE</i> (ΔM)	<i>d</i>	<i>Power</i>
Full data	2.02*	204.98	0.045	2.47	1.23	0.25	0.926
Listwise deletion	2.08*	194.88	0.039	2.62	1.26	0.26	0.915
Mean imputation	2.19*	204.66	0.030	2.62	1.20	0.27	0.926
Regression imputation	1.84	204.08	0.067	2.25	1.22	0.23	0.926
EM imputation	2.09*	204.64	0.038	2.50	1.20	0.26	0.926
Multiple imputation	1.87*	204.08	0.038	2.59	1.24	0.21	0.926

Note. *n* = 207. *df* = degrees of freedom. The *t* and *df* values are reported after adjustment for unequal sample sizes and unequal group variances. ΔM = mean difference. *d* = Cohen's *d*; * *p* < .05; ** *p* < .01; *** *p* < .001

Table 5. Independent samples *t* test results for education attainment with 10% missing data under various missing data handling methods using the 2000 census data

	Summary statistics					
	Virginia			Wisconsin		
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Full data	135	19.70	11.47	72	17.22	6.16
Listwise deletion	128	19.87	11.38	63	17.28	6.23
Mean imputation	135	19.87	11.08	72	17.28	5.82
Regression imputation	135	19.67	11.17	72	16.83	6.10
EM imputation	135	19.83	11.08	72	17.48	5.84
Multiple imputation	135	19.76	11.35	72	17.37	6.64

	<i>t</i> test statistics						
	<i>t</i>	<i>df</i>	<i>p</i>	ΔM	<i>SE</i> (ΔM)	<i>d</i>	<i>Power</i>
Full data	2.02*	204.98	0.045	2.47	1.23	0.25	0.926
Listwise deletion	1.82	183.57	0.071	2.37	1.31	0.23	0.894
Mean imputation	2.02*	204.98	0.044	2.37	1.17	0.25	0.926
Regression imputation	2.18*	204.96	0.031	2.63	1.21	0.27	0.926
EM imputation	1.79	204.99	0.075	2.11	1.18	0.22	0.926
Multiple imputation	1.87	202.46	0.071	2.37	1.27	0.21	0.926

Note. *n* = 207. *df* = degrees of freedom. The *t* and *df* values are reported after adjustment for unequal sample sizes and unequal group variances. ΔM = mean difference. *d* = Cohen's *d*; * *p* < .05; ** *p* < .01; *** *p* < .001

MISSING DATA GUIDELINES

($t = 1.79$, $p = .075$), and multiple imputation ($t = 1.87$, $p = .071$). These contradictory results stand in sharp contrast to results of tests of hypothesis obtained earlier in example 1 and underscore the risks inherent in using any missing data handling method when a large proportion of data is missing in a small sample.

Discussion

The primary objective of this study was to formulate general guidelines that can assist educational researchers in the selection of appropriate missing data handling methods under various combinations of sample size, proportion of missing data, and analytical method. By keeping all of these factors constant, any observed differences in performance of missing data handling methods can more or less be attributed to the relative efficiency of those methods. The statistical analyses conducted in this study can be thought of as a response to recommendations made in earlier studies such as Roth (1994) and Young et al. (2011) who identified a need for guidelines that can help researchers choose missing data handling methods under a variety of scenarios.

Although previous research exists that has looked at the effect of factors such as sample size, proportion of missing data, and method of analysis on the effectiveness of missing data handling methods, there are no clear cut guidelines which can inform a researcher as to which missing data handling method is best under which circumstances. Prior studies used different samples with varying proportions of missing data under different analytical methods which makes it very difficult to isolate the effect of any single factor. The present study is an attempt to rectify this state of affairs. It is hoped that insights provided by the findings of this study will further publicize the issues involved and encourage further research in this direction.

In some respects the present study has been able to confirm and support earlier findings. For example, our statistical results imply that listwise deletion is one of the simplest, easily justified, and least computation-intensive methods under large sample and low missing data conditions when the objective is to obtain consistent and unbiased estimates of population parameters (Haitovsky, 1968; Wayman, 2003; Young et al., 2011). On the other hand, the use of this method comes at the price of sacrificing additional statistical power that can be gained by imputing missing data. One can make a case that if sample size is large enough such that achievement of adequate power is not a concern, then listwise deletion provides one of the least risky (since it avoids adding another layer of measurement error to the data) and most quickly deployable missing data handling methods. Even in

cases where listwise deletion is not the best missing data handling method, for instance in terms of efficiency, it still remains an attractive choice because the efficiency gains offered by competing methods are often trivial making it difficult to justify the increased computational complexity in statistical analyses due to their employment.

We further confirmed the general finding of past studies that if missing data imputation is unavoidable, then the two best methods for such imputation are maximum likelihood imputation (e.g. Expectation-maximization imputation) and multiple imputation (Graham et al., 1996; Wayman, 2003; Peugh and Enders, 2004; Peng et al., 2006; Young et al., 2011; Knol et al., 2010). This can be clearly seen from the figures presented in Table 1 which show that ML and multiple imputation methods performed best in 22 out of 24 (91.6%) scenarios depicted therein. In order to get a more complete ranking of the five missing data handling methods used in this study, we used a simple scoring method where the least-performing to best-performing methods received a score from 1 to 5 for each of the 120 possibilities based on sample size (small, medium, large), proportion of missing data (low, high), the five missing data handling methods, and the four methods of analysis. The sum of scores across missing data handling methods revealed the following ranking and total scores: multiple imputation, 104; expectation maximization, 83; listwise deletion, 65; regression imputation, 63; and mean imputation, 45.

Although listwise deletion is in the third place in this ranking we reiterate our earlier contention that it is often preferable over other methods when the gain in estimation accuracy offered by those methods is trivial. This ranking of missing data handling methods also makes intuitive sense as it ranks these methods in the order of their mathematical sophistication, ranging from the most sophisticated, multiple imputation which offers most realistic modeling of random variation, to the least sophisticated, mean imputation method that offers no accommodation for random variability.

The important thing to note here is that the positive effect of gain in accuracy of parameter estimates due to missing data imputation does not always dominate the negative effect of measurement error introduced by such imputation. For instance, our results showed that in many instances listwise deletion, that is the no imputation method, worked better than some imputation methods but not others even after controlling for method of analysis, sample size, and proportion of missing data. For example, in our simulation two-way ANOVA for a medium sample with high proportion of missing data, listwise deletion performed better than mean imputation but worse than multiple imputation. For mean imputation in this scenario the positive effect of missing data imputation was dominated by the

MISSING DATA GUIDELINES

negative effect of larger measurement error due to that imputation. On the other hand, the reverse was true for multiple imputation where the positive effect of missing data imputation dominated the negative effect of larger measurement error due to such imputation. The message here is that missing data imputation is not always an improvement over non-imputation and that some missing data imputation methods can actually cause more harm than benefit.

An important implication of our statistical results is that missing data imputation can be beneficial in raising the statistical power of tests of hypothesis. In our simulated data relative power gain ranged between 0% and 28.8% while absolute power gain ranged between 0 and .12, depending on sample size, proportion of missing data, and method of analysis used. The gains in statistical power were pronounced for small samples, $n \leq 50$, in general (min gain = .003 or 0.4%; max gain = .11 or 28.8%; mean gain = .03 or 10.4%) and for small samples with high proportions of missing data ($m > 5\%$) in particular (min gain = .003 or 2.87%; max gain = .11 or 28.8%; mean gain = .04 or 14.9%). For sample sizes exceeding 200, statistical power was not an issue for any of the four methods of analysis adopted in this study (min power = .98; max gain = .01 or 1.2%). Similarly the gains in power were modest when proportion of missing data was 5% or less (max gain = .03 or 6.7%). The bottom line here is that statistical power by itself can be an important consideration for choosing missing data imputation even in cases where the non-missing pre-imputation data represents the target population well and listwise deletion is a viable option. This is especially true for small samples with large proportions of missing data.

The importance of statistical power issues highlighted in the preceding paragraphs should not be taken to mean that population representation is a minor consideration. Even when sample size is large and statistical power is not an issue, the occurrence of missing data can transform the sample in such a way that it is no longer representative of its target population. In such cases it is important to impute missing data or alternately, if possible, to use adjusted sampling weights in order to make the sample representative again. One may argue that the use of sampling weights is preferable over missing data imputation because the former method does not introduce additional measurement error.

Recommendations Based on Sample Size

Regarding choice of missing data handling method our general recommendation is that if (1) sample size is large enough for adequate power, and (2) sample is representative of the target population, then use listwise deletion. In cases where either of these conditions is not met the best methods are multiple imputation and maximum likelihood imputation. It is important to note here that these recommendations are for missing data that are either missing at random (MAR) or missing completely at random (MCAR), and not for data that are not missing at random (NMAR).

When sample size is large, $n \geq 1,000$, lack of statistical power is generally not an issue as clearly demonstrated by our simulated results and empirical data examples. The decision to impute missing data thus depends on whether or not the non-missing data are still representative of the target population. For small samples, in terms of gain in accuracy of estimation, the best available methods of missing data imputation are maximum likelihood imputation and multiple imputation. Although strictly speaking multiple imputation on average performs better than ML imputation in small samples we recommend using more than one imputation method in general when the sample size is small and in particular when sample size is small and proportion of missing data is high in order to lower the risk of getting into the unfortunate situation where the negative effect of an increase in measurement error due to imputation exceeds the positive effect of a gain in estimation accuracy due to that imputation.

Our recommendations for choice of missing data handling method are summarized in [Figure 2](#). If the missing data are MCAR and the resulting sample after listwise deletion provides adequate power for tests of hypotheses, then listwise deletion should be used. If the missing data are MAR, then listwise deletion should only be used if the resulting sample after listwise deletion is still representative of the population and there is adequate power for tests of hypotheses. Finally, if missing data is NMAR, then the missing data mechanism must be modeled as part of the estimation process. Because the term NMAR is an umbrella term for all sorts of non-random missing data mechanisms, the exact modeling process depends on the type of non-randomness present in the missing data. For example, if the missingness is due to selection bias, Heckman correction can be used ([Heckman, 1979](#)). We recommend multiple imputation and maximum likelihood imputation as the methods of choice.

MISSING DATA GUIDELINES

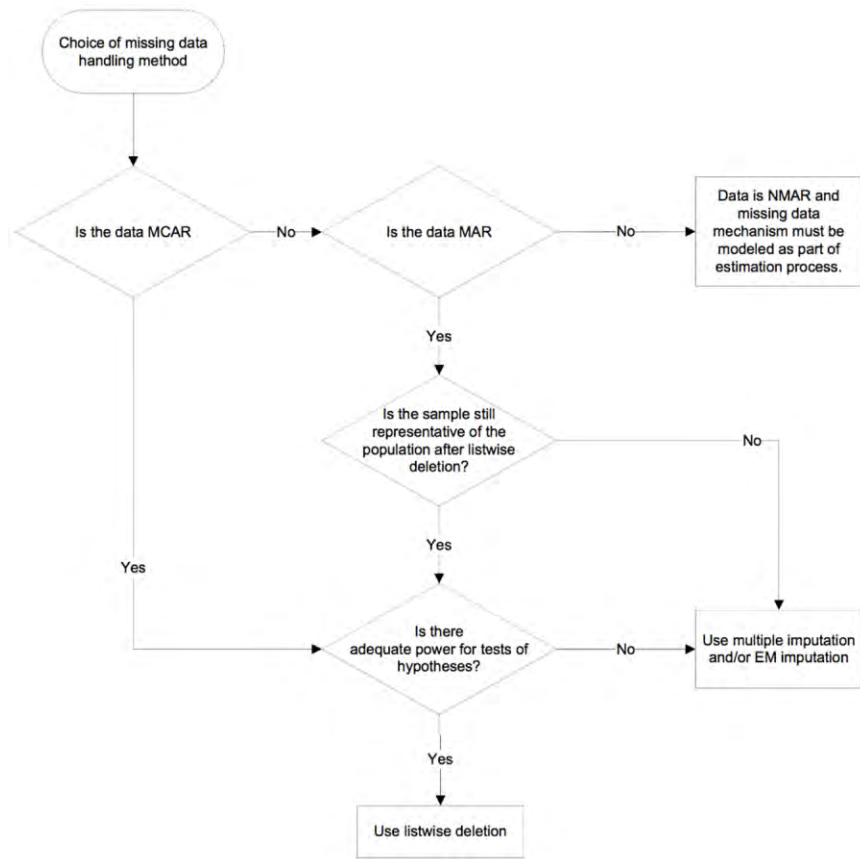


Figure 2. The decision process governing choice of missing data handling method

Scope for Future Research

There are several directions for future research. First, more work needs to be done on the effect of missing data handling methods on method of analysis. All four methods of analysis adopted for statistical analyses presented in this study, one sample t test, independent samples t test, two-way ANOVA, and multiple regression, are special cases of the general linear model. It would be interesting to see whether the guidelines developed here are also applicable to nonlinear models, for example models of count data such as logistic regression. There is also further scope for testing these guidelines in context of longitudinal, repeated measures, and multi-level models.

The second potential line of research is to focus on application. Future studies can take an applied approach and use real-life datasets from various subfields of

education in order to evaluate the effectiveness of guidelines presented in this study. The importance of simulation work notwithstanding, it is the presence or lack of empirical evidence which is most important in determining whether or not such guidelines may see widespread acceptance in educational research.

References

- Acock, A. (2005). Working with missing values. *Journal of Marriage and Family*, 67(4), 1012-1028. doi:10.1111/j.1741-3737.2005.00191.x
- Afifi, A., & Elashoff, R. (1966). Missing observations in multivariate statistics: I. Review of the Literature. *Journal of the American Statistical Association*, 61(315), 595-604. doi:10.2307/2282773
- Allison, P. (2001). *Missing data*. (Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-136). Thousand Oaks, CA: Sage. doi:10.4135/9781412985079
- Alosh, M. (2009). The impact of missing data in a generalized integer-valued autoregression model for count data. *Journal of Biopharmaceutical Statistics*, 19, 1039–1054. doi:10.1080/10543400903242787
- Andridge, R., & Little, R. (2010). A review of hot deck imputation for survey non-response. *International Statistical Review*, 78(1), 40-64. doi:10.1111/j.1751-5823.2010.00103.x
- Brick, J., & Kalton, J. (1996). Handling missing data in survey research. *Statistical Methods in Medical Research*, 5, 215-238. doi:10.1177/096228029600500302
- Cohen, J. (1992). Quantitative methods in Psychology: A power primer. *Psychological Bulletin*, 112(1), 155-159. doi:10.1037/0033-2909.112.1.155
- Graham, J., Hofer, S., & MacKinnon, D. (1996). Maximizing the usefulness of data obtained with planned missing value patterns: An application of maximum likelihood procedures. *Multivariate Behavioral Research*, 31(2), 197-218. doi:10.1207/s15327906mbr3102_3
- Groves, R., Dillman, D., Eltinge, J., & Little, R. (2002). *Survey Nonresponse*. New York, NY: John Wiley & Sons, Inc.
- Groves, R., Fowler, F., Couper, M., Lepkowski, J., Singer, E., & Tourangeau, R. (2004). *Survey methodology*. Hoboken, NJ: John Wiley & Sons, Inc.

MISSING DATA GUIDELINES

- Haitovsky, Y. (1968). Missing data in regression analysis. *Journal of the Royal Statistical Society, Series B (Methodological)*, 30(1), 67-82.
- Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1), 153-161. doi:10.2307/1912352
- Jones, M. (1996). Indicator and stratification methods for missing explanatory variables in multiple linear regression. *Journal of the American Statistical Association*, 91(433), 222-230. doi:10.2307/2291399
- Knol, M., Janssen, K., Donders, A., Egberts, A., Heerdink, E., Grobbee, D., Moons, K., & Geerlings, M. (2010). Unpredictable bias when using the missing indicator method or complete case analysis for missing confounder values: an empirical example. *Journal of Clinical Epidemiology*, 63, 728-736. doi:10.1016/j.jclinepi.2009.08.028
- Madow, W., Nisselson, H., & Olkin, I. (Eds.). (1983). *Incomplete data in sample surveys, Volume 1: Report and case studies*. New York, NY: Academic Press.
- Madow, W., & Olkin, I. (Eds.). (1983). *Incomplete data in sample surveys, Volume 3: Proceedings of the symposium*. New York, NY: Academic Press.
- Madow, W., Olkin, I., & Rubin, D. (Eds.). (1983). *Incomplete data in sample surveys, Volume 2: Theory and bibliographies*. New York, NY: Academic Press.
- National Center for Education Statistics. (2003). *Program for International Student Assessment* [Data file]. Retrieved from <http://nces.ed.gov/surveys/pisa/datafiles.asp>
- Peng, C., Harwell, M., Liou, S., & Ehman, L. (2006). Advances in missing data methods and implications for educational research. In S. S. Sawilowsky (Ed.), *Real Data Analysis* (pp. 31-78). Charlotte, NC: New Information Age Publishing.
- Peugh, J., & Enders, C. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, 74(4), 525-556. doi:10.3102/00346543074004525
- Raymond, M. & Roberts, D. (1987). A comparison of methods for treating data in selection research. *Educational and Psychological Measurement*, 47(1), 13-26. doi:10.1177/0013164487471002
- Roth, P. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology*, 47(3), 537-560. doi:10.1111/j.1744-6570.1994.tb01736.x

- Rubin, D. (1987). *Multiple imputation for nonresponse in surveys*. New York, NY: John Wiley & Sons, Inc. doi:10.1002/9780470316696
- Schafer, J., & Graham, J. (2002). Missing data: Our view on the state of the art. *Psychological Methods*, 7(2) 147-177. doi:10.1037//1082-989X.7.2.147
- U.S. Census Bureau. (2000). *Census of population and housing* [Data file]. Retrieved from <http://www.ers.usda.gov/data/education>
- Wayman, J. (2003, April). *Multiple imputation for missing data: What is it and how can I use it?* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.
- Young, W., Weckman, G., & Holland, W. (2011). A survey of methodologies for the treatment of missing values within datasets: limitations and benefits. *Theoretical Issues in Ergonomics Science*, 12(1), 15-43. doi:10.1080/14639220903470205

Conover's F Test as an Alternative to Durbin's Test

D. J. Best

University of Newcastle
Newcastle, Australia

J. C. W. Rayner

University of Newcastle
Newcastle, Australia

Data consisting of ranks within blocks are considered for balanced incomplete block designs. An F test statistic from ANOVA is better approximated by an F distribution than the Durbin statistic is approximated by a chi-squared distribution. Indicative powers demonstrate that the F test is generally superior to Durbin's test.

Keywords: Analysis of variance, balanced incomplete blocks, F tests, powers, ranks data, taste-tests, test sizes

Introduction

Sometimes the number of treatments to be compared is so large that a complete blocks experiment is impractical. This happens, for example, in some agronomic variety trials. A balanced incomplete block (BIB) design can be used in such a situation. In sensory evaluation trials loss of sensitivity can occur when the subjects are not be able to compare more than a few products with any certainty. Again BIB designs are useful.

Consider a balanced incomplete block design with the data being ranks within blocks. A traditional test for treatment differences for such a scenario is the Durbin (1951) test, based on the statistic D , given by

$$D = c \sum_{i=1}^t \left(\bar{R}_i - \frac{k+1}{2} \right)^2$$

Dr. Best is a conjoint Associate Professor in the School of Mathematical and Physical Sciences. Email him at john.best@newcastle.edu.au. Dr. Rayner is a conjoint Professor in the School of Mathematical and Physical Sciences. Email him at john.rayner@newcastle.edu.au.

in which there are t treatments, k of which are ranked in each of b blocks. If r_{ij} is the rank given to treatment i on block j then \bar{R}_i is defined as the mean rank over blocks for the i^{th} treatment. For untied data $c = 12(t-1)r/\{t(k^2-1)\}$ where each treatment is ranked r times, with $r < b$. For tied data, if $V = \sum_{i,j} r_{ij}^2 / (rt) - (k+1)^2 / 4$ then $c = (t-1)r/(tV)$.

It is well known that D has an asymptotic χ_{t-1}^2 distribution. However for values of (t, b, k, r) met in practice this approximation to the distribution of D can be poor. See, for example, Fawcett and Salter (1987). This has led to the suggestion to use a permutation test to obtain p-values for D . See, for example, Bi (2009) and Higgins (2004) who does not consider BIBs but who generally advocates permutation tests. However, software to calculate a p-value for D via a permutation test may not always be readily available. A scientist without access to software for a permutation test based on D might find carrying out a permutation test a challenge. Conover (1999, p. 389) suggests carrying out an analysis of variance (ANOVA) on the ranks and using the F test for treatment differences. This F test is based on ‘adjusted’ sums of squares from the general linear model readily available in statistical packages such as JMP (use ‘fit model’) and MINITAB (use ‘glm’). These packages also readily give appropriate multiple comparisons.

Literature reviews did not reveal any previously published small sample studies examining the validity of this F test approach, although Conover (1999, p. 390) suggested it improves on the Durbin test. As above we observe that we are considering situations where the raw data are ranks or ranks of ordered categorical data. Many studies have compared parametric and nonparametric tests when data are continuous measurements. See, for example, Kelley and Sawilowsky (1997) and the references therein. However, such studies are not the focus of our article.

Sizes and powers

Test sizes based on 100,000 samples for each (t, b, k, r) combination in Table 1 were carried out to check the suitability of the χ_{t-1}^2 and $F_{t-1,df}$ distributions for obtaining p-values. Note that, as usual, $df = bk - t - b + 1$. Table 1 sizes are for data with untied ranks with χ_{t-1}^2 and $F_{t-1,df}$ critical values and were found using permutation tests. Sizes for D in this study agree with those of Fawcett and Salter (1987). The sizes for the F test statistic F improve on those for D based on the χ_{t-1}^2 approximation and indicate the F distribution can be used to obtain p-values for F . If carrying out a permutation test is not convenient the F probabilities are a considerable improvement over the χ^2 probabilities for all bk .

AN ALTERNATIVE TO DURBIN'S TEST

Table 1. Actual test sizes for nominal 5% level test for various BIBs with no ties

(t, b, k, r)	D	F	bk
(4, 6, 2, 3)	0.000	0.000	12
(4, 4, 3, 3)	0.000	0.073	12
(5, 10, 2, 4)	0.000	0.116	20
(5, 5, 4, 4)	0.026	0.062	20
(5, 10, 3, 6)	0.035	0.061	30
(6, 15, 2, 5)	0.022	0.050	30
(6, 10, 3, 5)	0.026	0.062	30
(6, 15, 4, 10)	0.030	0.051	60
(6, 20, 3, 10)	0.040	0.056	60
(7, 7, 3, 3)	0.000	0.087	21
(7, 7, 4, 4)	0.025	0.055	28
(7, 21, 2, 6)	0.017	0.069	42

Table 2. Actual test sizes for nominal 5% level test for various BIBs with ties

(t, b, k, r)	D	F	bk
(4, 6, 2, 3)	0.000	0.006	12
(4, 4, 3, 3)	0.005	0.051	12
(5, 10, 2, 4)	0.001	0.045	20
(5, 5, 4, 4)	0.018	0.044	20
(5, 10, 3, 6)	0.031	0.052	30
(6, 15, 2, 5)	0.012	0.053	30
(6, 10, 3, 5)	0.024	0.050	30
(6, 15, 4, 10)	0.042	0.050	60
(6, 20, 3, 10)	0.040	0.050	60
(7, 7, 3, 3)	0.001	0.054	21
(7, 7, 4, 4)	0.020	0.050	28
(7, 21, 2, 6)	0.020	0.054	42

To allow for ties, sizes were calculated as in Brockhoff et al. (2004, section 4 and also see the discussion in section 6). For each block and treatment one of the scores $1, 2, \dots, k$ was randomly assigned, each with probability $1/k$. These values were ranked by block with ties given mid-rank values. This was repeated 100,000 times for each of the (t, b, k, r) combinations in Table 2. Very infrequently the value V or the error sum of squares was zero. Such data sets were discarded and new ones inserted. Sizes for D are still poor but those for F are better for the ties case than for the no ties case. If, for tied ranks, permutation tests rather than the Monte Carlo tests suggested herein had been used to get sizes for Table 2, results would have been conditional on a ties structure and so not of as general applicability as those given.

A power comparison between the tests is now provided based on D and F . In practice it is expected that most scientists will use the χ^2 and F distributions and so powers based on these are provided. However Tables 1 and 2 show that the test based on D hardly ever has size near 0.05; thus, the D powers based on χ^2 critical points will be disadvantaged in comparison to the F powers based on the F distribution. If test sizes for tied data are examined, it is observed that D sizes for $(t, b, k, r) = (6, 15, 4, 10)$ and $(6, 20, 3, 10)$ are at least not too far from 0.05 and so the D test should not be too disadvantaged. Subsequent powers are calculated following the size method but with all treatments in a given treatment group having probabilities (p_1, p_2, p_3, p_4) of getting a score $(1, 2, 3, 4)$ respectively in any given block instead of $(0.25, 0.25, 0.25, 0.25)$. Thus for $(t, b, k, r) = (6, 15, 4, 10)$ with probabilities $(0.25, 0.25, 0.25, 0.25)$ for treatments 1, 2 and 3, and probabilities $(0.08, 0.08, 0.42, 0.42)$ for treatments 4, 5 and 6 for a nominal 5% level of significance, it was found that the D and F test powers are 0.31 and 0.34 respectively. Recall that under the null hypothesis all treatment probabilities are $(0.25, 0.25, 0.25, 0.25)$. The powers here are close, and the difference could be explained by the discrepancy in the actual sizes. It is expected that - if there was no difference in the sizes - the sizes would be, as here, very close, and there would be no reason, based on power, to use D rather than F . For $(t, b, k, r) = (6, 20, 3, 10)$ and treatment group probabilities as above, then the D and F powers are respectively 0.26 and 0.29. Again, these are very similar and any difference may well be due to the size advantage enjoyed by the F test. It must be stressed that the powers just given are for a BIB design where the actual size was near the nominal size. For the many BIB designs where this is not so, powers of the test based on D would be very poor compared to those of the test based on F .

To further compare the powers of the tests based on D and F and to check whether or not it is the slight size difference that is causing the differences in power, Table 3 gives powers for $(t, b, k, r) = (6, 15, 4, 10)$ for a number of alternative treatment probabilities, using an estimated critical value of 10.64 for the test based on D . Also given are powers using the χ^2_5 critical value of 11.07, which gives a test size of 0.042, whereas 10.64 gives a test size of 0.05.

In all cases the F test power is found to be slightly superior to the Durbin test power; using the estimated critical value of 10.64 it is superior by so little as to be inconsequential. Using the χ^2_5 critical value the difference is small but not inconsequential. Therefore, use of the F test is recommended based on its test sizes being closer to nominal than the Durbin test sizes. Moreover the F test power is generally not inferior, and when the Durbin test has a low size, it is generally

AN ALTERNATIVE TO DURBIN'S TEST

inferior. The F test is easy to use and has ready availability of multiple comparisons in general linear model platforms.

Table 3. Powers for a 5% significance level, $(t, b, k, r) = (6, 15, 4, 10)$, with ties allowed and alternative probabilities as shown

Treatment groups	Alternative probabilities	$D (11.07)$	$D (10.64)$	F
(1, 2, 3) (4, 5, 6)	(0.25, 0.25, 0.25, 0.25) (0.15, 0.15, 0.35, 0.35)	0.112	0.129	0.133
(1, 2, 3) (4, 5, 6)	(0.25, 0.25, 0.25, 0.25) (0.1, 0.1, 0.4, 0.4)	0.229	0.258	0.262
(1, 2, 3) (4, 5, 6)	(0.25, 0.25, 0.25, 0.25) (0.05, 0.05, 0.45, 0.45)	0.437	0.473	0.478
(1, 2, 3) (4, 5, 6)	(0.25, 0.25, 0.25, 0.25) (0.02, 0.02, 0.48, 0.48)	0.598	0.634	0.639
(1, 2) (3, 4, 5, 6)	(0.25, 0.25, 0.25, 0.25) (0.15, 0.15, 0.35, 0.35)	0.102	0.119	0.122
(1, 2) (3, 4, 5, 6)	(0.25, 0.25, 0.25, 0.25) (0.1, 0.1, 0.4, 0.4)	0.207	0.233	0.236
(1, 2) (3, 4, 5, 6)	(0.25, 0.25, 0.25, 0.25) (0.05, 0.05, 0.45, 0.45)	0.400	0.433	0.437
(1, 2) (3, 4, 5, 6)	(0.25, 0.25, 0.25, 0.25) (0.02, 0.02, 0.48, 0.48)	0.546	0.578	0.584
(1, 2) (3, 4) (5, 6)	(0.25, 0.25, 0.25, 0.25) (0.1, 0.1, 0.4, 0.4) (0.02, 0.02, 0.48, 0.48)	0.551	0.584	0.588
(1, 2) (3, 4) (5, 6)	(0.25, 0.25, 0.25, 0.25) (0.1, 0.1, 0.4, 0.4) (0.15, 0.15, 0.35, 0.35)	0.175	0.197	0.201

Examples

Ice cream data

Suppose, as in Conover (1999, p. 390) that seven varieties of ice cream are to be compared. Also suppose it is known that tasting more than three ice creams at a time will result in poor responses due to sensory fatigue. The seven ice cream judges are each asked to rank three of the seven varieties. The results are in Table 4.

Table 4 shows that $t = b = 7$, $r = k = 3$ and each variety is compared with every other variety once. This is a balanced incomplete block layout; no ties are observed

and $D = 12$. Using the χ^2_6 approximation the p-value is 0.06 and so with a 5% level of significance it may be concluded that no difference exists in the preference for the seven varieties. However Conover calculates an exact p-value of 0.018. This study calculated $F = 8$ and, using the $F_{6,8}$ distribution, the p-value is 0.005: this is much closer to the exact Conover p-value for D , is easier to calculate and is significant at the 5% level. Knowing that the χ^2 approximation to D is poor, it is necessary to reverse the initial judgement and conclude that varieties are not equally preferred. As here the χ^2 p-values are often too conservative.

Table 4. Rankings of seven ice cream varieties

Judge	Variety						
	1	2	3	4	5	6	7
1	2	3	-	1	-	-	-
2	-	3	1	-	2	-	-
3	-	-	2	1	-	3	-
4	-	-	-	1	2	-	3
5	3	-	-	-	1	2	-
6	-	3	-	-	-	1	2
7	3	-	1	-	-	-	2
Sum	8	9	4	3	5	6	7

Table 5. Rankings for breakfast cereals

Judge	Cereal				
	A	B	C	D	E
1	1.5	1.5	3	-	-
2	1	2.5	-	2.5	-
3	1.5	3	-	-	1.5
4	1	-	2	3	-
5	1.5	-	3	-	1.5
6	1	-	-	3	2
7	-	2	3	1	-
8	-	2.5	2.5	-	1
9	-	3	-	2	1
10	-	-	3	2	1
Sum	7.5	14.5	16.5	13.5	8

Breakfast cereal (tied) data

Kutner et al. (2005, section 28.1) consider a taste-test in which five breakfast cereals ($t = 5$) were scored on a ten point hedonic scale by ten judges ($b = 10$) three at a time ($k = 3$). Each cereal was tasted six times ($r = 6$). The ranked data are shown in Table 5 (note that there are tied ranks).

It was found that $D = 14.92$ with a χ^2_4 p-value of 0.005 and $F = 11.56$ with an $F_{4,16}$ p-value of 0.0001. Using a 5% significance level a decision would be made that there was a difference in the preference ranking of the cereals.

Conclusion

The test based on the ANOVA F statistic F provides an easily applied alternative to Durbin's rank test. The test based on the F statistic has better test sizes than the test based on D , has better power if chi-squared critical values are used for D , and can be calculated using the general linear model software available in many statistical packages, which also readily provide multiple comparisons. Based on the results in this study, it is suggested that, for $bk \geq 50$, the F statistic p-value based on the F distribution can be used rather than p-values from permutation or Monte Carlo tests. For smaller bk the F probabilities are a considerable improvement over the χ^2 probabilities and should be used when carrying out a permutation test is not convenient.

References

- Bi, J. (2009). Computer intensive methods for sensory data analysis, exemplified by Durbin's rank test. *Food Quality and Preference*, 20, 195-202.
- Brockhoff, P. B., Best, D. J. and Rayner, J. C. W. (2004). Partitioning Anderson's statistic for tied data. *Journal of Statistical Planning and Inference*, 121(1), 93-111.
- Conover, W. J. (1999). *Practical nonparametric statistics* (3rd ed.). New York: Wiley.
- Durbin, J. (1951). Incomplete blocks in ranking experiments. *British Journal of Psychology* (Statistical Section), 4, 85-90.
- Fawcett, R. and Salter, K. (1987). Distributional studies and the computer: an analysis of Durbin's rank test. *The American Statistician*, 41, 81-83.

Higgins, J. J. (2004). *Introduction to Modern Nonparametric Statistics*. Belmont, CA: Duxbury Press.

Kelley, D. L. and Sawilowsky, S. S. (1997). Nonparametric alternatives to the F statistic in analysis of variance. *Journal of Statistical Computing and Simulation*, 58, 343-359.

Kutner, M., Nachtsheim, C., Neter, J. and Li, W. (2005). *Applied linear statistical models* (5th ed.). Boston: McGraw-Hill Irwin.

A Comparison of Methods for Group Prediction with High Dimensional Data

Holmes Finch

Ball State University
Muncie, IN

High dimensional data is the situation in which the number of variables included in an analysis approaches or exceeds the sample size. In the context of group classification, researchers are typically interested in finding a model that can be used to correctly place an individual into their appropriate group; e.g. correctly diagnose individuals with depression. However, when the size of the training sample is small and the number of predictors used to differentiate the groups is larger, standard approaches such as discriminant analysis may not work well. In order to address this issue, statisticians have developed a number of tools designed for supervised classification with high dimensional data. The goal of this simulation study was to compare several such approaches for supervised classification with high dimensional data in terms of their ability to correctly classify individuals into groups, and to identify the number of variables associated with group separation. Results of the study showed that the Random Forest ensemble recursive partitioning algorithm was optimal for group prediction, while the Nearest Shrunken Centroid and Regularized Discriminant Analysis methods were optimal for identifying the number of salient predictor variables. The standard linear discriminant analysis approach was generally the worst performer across all high dimensional simulated conditions. Implications of these results to practice and directions for future research are discussed.

Keywords: Group prediction, Discriminant Analysis, High dimensional data, Regularization methods

Introduction

High dimensional data refers to the case where the number of variables to be included in an analysis is equal to or exceeds the sample size (Bühlmann & van de Geer, 2011), and is written symbolically as $p \gg n$. High dimensional data can create a variety of problems for many standard data analytic techniques, including those used in prediction and classification. In particular, when the number of predictors exceeds the sample size it is frequently not possible to obtain model parameter

Holmes Finch is the George and Frances Ball Distinguished Professor of Educational Psychology, and a professor of statistics and psychometrics. Email at whfinch@bsu.edu.

estimates because covariance matrices are often singular. In addition, in the high dimensional case there may be more unknown parameters than known data, leading to indeterminate estimation problems. Finally, in the high dimensional data case the correlations among variables is often very high, making parameter estimation very difficult. The result of all of these problems is that parameters and their corresponding standard errors are frequently not estimable in the high dimensional case. Furthermore, any estimates that are obtained are likely to be ill-conditioned and therefore unreliable (Kriegel, Kröger, & Zimek, 2009).

Given these difficulties, researchers have developed a set of statistical methods for the problem of high dimensional data. These methods are useful in a variety of contexts, including fitting of linear models, clustering observations based on a number of variables (often referred to as features in high dimensional literature), and classification of individuals into one of several groups, using many predictors. The focus of the current research is on this latter application to group classification. Often in standard data problems where $n > p$, such classification is done using discriminant analysis. However, as we will see below, this approach is ill suited for use when $p \gg n$, or even when p approaches n (Hastie, Buja, & Tibshirani, 1995). This Monte Carlo simulation study examines several methods that have been proposed for the high dimensional classification problem, including several based on discriminant analysis, as well as a variation of nearest centroid classification, and the recursive partitioning random forest methodology. The remainder of the manuscript is organized as follows: First we discuss discriminant analysis and explain why its use when $p \gg n$ is problematic, followed by a description of alternative classifiers that have been proposed for this case. After next describing the study goals, we then outline the simulation study design, followed by a description of the simulation results. Finally, we discuss the results of our simulation and place them in the context of the broader high dimensional classification literature.

Goals of the current study

The goal of the current study was to compare the performance of several methods for group classification in the presence of high dimensional data. This comparison was made using both simulated data and an existing data example. This work adds to the literature in the field in three primary ways. First of all, there has not been another published study in which all of these classification methods have been compared with one another using Monte Carlo methods. While prior research has demonstrated the utility of several of these approaches using extant data (e.g.

Clemmensen, Hastie, Witten, & Ersbøll, 2011; Zhang, Dai, Xu, & Jordan, 2010; Hastie, Buja, & Tibshirani, 1995) or small simulations (Hastie, Tibshirani, & Friedman, 2009), no prior published study has systematically compared the performance of all of the methods described here, each of which has been suggested as a possible alternative for the high dimensional case. Thus, the current study should further the literature in this regard by providing researchers with more information about which such tools might be optimal under which conditions. The second goal of this study was to investigate the performance of RF in the presence of high dimensional data. While there have been some initial calls for such simulation research (Xu, Huang, Williams, Wang, & Ye, 2012), and some demonstrations with existing data (e.g. Zhang, Yu, Singer, & Xiong, 2001) there has not been a great deal of work done examining RF in this context. Finally, the third goal of this study was to introduce methods for high dimensional group prediction to social science researchers, in particular. Traditionally these methods have been used primarily with gene expression data, as a review of Bühlmann and van d Geer (2011) demonstrates. However, there are scenarios in the social sciences in which researchers are faced with small samples as well (e.g., Siklos & Kerns, 2007; Palmer, 2006; Sanden, 2008).

Methods for classification with high dimensional data

Linear Discriminant Analysis

Perhaps one of the most widely used classification methods is linear discriminant analysis (LDA). This technique, which is based upon a multivariate linear model, is used when there exists a grouping variable and a set of predictors that are believed to distinguish members of the 2 or more groups. The algorithm identifies weights for each of the predictors such that their linear combination maximally separates the groups from one another (Huberty & Olejnik, 2006). This linear combination appears in equation (1).

$$C_{ji} = \beta_{j0} + \sum \beta_{jm} x_{mi} + \ln \left(\frac{n_j}{N} \right) \quad (1)$$

The terms in (1) are defined as:

C_{ji} = Classification score for group j for subject i

β_{j0} = Constant for group j

β_{jm} = Weight for predictor m for group j

x_{mi} = Value of predictor m for subject i

n_j = Sample size for group j

N = Total sample size

The natural log of the ratio of group size to the total sample serves as prior information about the relative frequency of the group in the population. In many applications, this prior probability of group membership is calculated using the values of n_j and N from the sample, as described above. However, the prior probability can also be provided directly by the researcher, bypassing the use of relative group size in the sample. This might be a useful strategy if it is known that the sample is not representative of the population in terms of the relative frequency with which members of each group appear. Determination of the coefficients in (1) is made so as to maximize the following criterion:

$$\beta_j^T \Sigma_b \beta_j \text{ subject to } \beta_j^T \Sigma_w \beta_j \leq 1 \quad (2)$$

Here, the coefficients (β) are as defined previously, with Σ_b being the between class covariance matrix, and Σ_w the within class covariance matrix. The resulting linear combination in (1) can be used to determine category membership for each observation in the original training data or in a cross-validation sample. Values of C_j are calculated for each member of the sample, and individuals are classified into the group for which they have the largest such score.

In terms of determining variable importance in terms of group classification, researchers typically rely on the structure matrix, which can be interpreted as the correlation matrix between the individual predictors and C . While there is not a hypothesis test for these values, recommendations for cut values have been suggested, including 0.32 (Tabachnick & Fidell, 2013), which will be used in the current study. Thus, absolute values of the structure matrix elements greater than 0.32 are taken as indicative that a predictor contributes to a classification solution.

In the case when $P > N$, which is the focus of this study, LDA is not typically a good choice for group classification because the within class covariance matrix

estimated using the sample is likely to be singular (Witten & Tibshirani, 2011). Even when this is not the case, Witten and Tibshirani have shown that the classifier in (1) will most likely exhibit a high variance, thereby degrading the resulting prediction model. Finally, in the case when P itself is large (irrespective of the size of N), LDA can prove problematic because the classifier in (1) will, by definition include all of the predictor variables, potentially leading to problems with interpreting the classifier function (Witten & Tibshirani). Therefore, while it is a popular and frequently used tool for researchers interested in classification, LDA may not be appropriate for cases in which the number of predictor variables is almost as large as, or larger than the number of subjects in the sample (Hastie, Tibshirani, & Friedman, 2009). Given these limitations, we will need to turn to alternative methods of classification better suited to the high dimensional problem.

Penalized LDA

One alternative for high dimensional classification is penalized LDA (PLDA), as proposed by Hastie, Buja, and Tibshirani (1995). PLDA is based upon a regularization of the discriminant function; i.e. a reduction in the number of predictors (sometimes referred to as features) used to develop the prediction algorithm. By limiting the number of predictors, the resultant discriminant function should not suffer from the problems associated with LDA when $P > N$. The key to this method working optimally is the use of an appropriate regularization strategy. PLDA shares the basic methodology described above for LDA, including the form of (1) for the prediction algorithm. However, (2) is adjusted to the following:

$$\beta_j^T \Sigma_b \beta_j - P(\beta_j) \text{ subject to } \beta_j^T \Sigma_w \beta_j \leq 1 \quad (3)$$

In (3), P is a penalty function designed to regularize the set of predictor variables, by reducing it to only those that are most salient in differentiating the groups.

Witten and Tibshirani (2011) describe the penalty function to be used in (3) as the PLDA-L1 algorithm appearing in (4):

$$\beta_j^T \Sigma_b \beta_j - \lambda \sum |\sigma_j \beta_j| \text{ subject to } \beta_j^T \Sigma_w \beta_j \leq 1 \quad (4)$$

In this case, σ_j is the within class standard deviation for predictor j , and λ is a tuning parameter that is set by the researcher. When λ is large, the relative

importance of individual predictors is reduced, and some will even go to 0, meaning that they do not contribute to the classification function in (1) at all. The value of λ is determined through the use of jackknife cross-validation, in which each member of the sample is removed in turn and various values of the tuning parameter are used with each jackknife sample. The optimal value is determined to be the one that minimizes classification error across the cross-validation samples. In addition, the inclusion of σ_j means that predictors with greater variation within classes will contribute less to the overall classification function than those with less such variability. Witten and Tibshirani assert that using PLDA-L1 will result in a function involving a subset of the predictors, and is most appropriate if the researcher desires a relatively sparse classifier function. It should be noted that Witten and Tibshirani also describe a second method for determining the penalty in (3), based on the fused Lasso method of regularization (Tibshirani, Saunders, Rosset, Zhu, & Knight, 2005). However, this approach was not employed in the current study because it assumes a linear ordering of the predictors, which was felt to be a limitation to its use in many applied settings.

Regularized Discriminant Analysis (RDA)

Guo, Hastie, and Tibshirani (2007) introduced an alternative to LDA in the high dimensional case that focuses on shrinking the within class covariance matrix (Σ_w) in the sample toward the diagonal matrix, through the use of a tuning parameter, γ . This shrunken version of Σ_w takes the following form

$$\Sigma_{w\gamma} = \gamma \Sigma_w + (1 - \gamma) \text{diag}(\Sigma_w) \quad (5)$$

The value of γ ranges between 0 and 1, where $\gamma \rightarrow 1$ corresponds to standard LDA, and $\gamma \rightarrow 0$ yields highly regularized discriminant functions, in which a small number of predictors contribute to the within class covariance matrix. As with PLDA, jackknife cross-validation is used to identify the optimal value of γ . When $\Sigma_{w\gamma}$ is obtained, it is applied to (2), and the classifier in (1) is developed based upon this shrunken within class covariance matrix. In practice, RDA yields a classification function that utilizes many fewer predictors than are present in the data, or than would be used in standard LDA, thus avoiding problems associated with complex classifiers containing many correlated predictors, as cited above.

Sparse Discriminant Analysis (SDA)

RDA regularized the set of predictors by applying a penalty to the within class covariance matrix. Alternatively, regularization could be achieved through applying the penalties directly to the discriminant function coefficients for the predictors (β_j). This approach, known as sparse discriminant analysis (SDA) was described in Clemmensen, Hastie, Witten, and Ersboll (2011). It is based on the elastic net (Zou & Hastie, 2005), which is used with linear models in the presence of high dimensional and/or highly collinear data. In the context of discriminant analysis, this elastic net approach seeks to minimize the following function:

$$Y\theta_j - X\beta_j^2 + \gamma\beta_j^T\Omega\beta_j + \lambda\beta_j \quad (6)$$

The parameters γ and λ are tuning parameters, Y is an indicator variable for whether an individual belongs to a particular group, θ_j is a score matrix, and Ω is a positive definite penalty matrix. In the current study, we use the elastic net approach suggested by Clemmensen, et al., such that $\Omega = \gamma I$, where I is the identity matrix. Jackknife cross-validation is used to determine the optimal values of γ and λ . The elastic net approach to regularization has some theoretical advantages over other approaches, including the lasso based PLDA method described above. Chief among these advantages are that highly correlated predictors tend to have similar coefficients in the final equation, and a greater number of predictors might be included in the final equation (Zou & Hastie, 2005).

Nearest Shrunken Centroids (NSC)

Another approach that we will examine for dealing with the high dimensionality classification problem is based upon an approach known as diagonal-covariance LDA. This method is based upon centroid classification, in which a multivariate mean (centroid) across all predictors is estimated for each group, and then new cases are placed in the class to whose centroid their scores are closest. NSC is a variant of this approach in which the class centroids are shrunken toward the overall centroid of the sample by an amount equal to a predetermined threshold value. The nearest centroid classification rule is expressed as:

$$\delta_{ij} = -\sum \frac{x_{im} - \bar{x}_{jm}}{s_m^2} + 2\ln\pi_j \quad (7)$$

In (7) δ_{ij} is the discriminant score of subject i for category j , \bar{x}_{jm} is the mean for predictor m in category j , s_m^2 is the variance of predictor m pooled across all categories, and x_{im} is the value of predictor m for subject i . An individual is placed into the group for which their value of δ_{ij} is smallest.

NSC adjusts the nearest centroid classification in the following way. First, the value d_{jm} is calculated, reflecting the difference between each group mean and the overall mean, as seen in (8).

$$d_{jm} = \frac{\bar{x}_{jm} - \bar{x}_m}{m_j(s_m + s_0)} \quad (8)$$

Here, terms are as defined in (7), with the addition that \bar{x}_m is the overall mean across categories for predictor m , $m_j^2 = \frac{1}{n_k} - \frac{1}{n}$, and s_0 is the median of the s_m values. Its purpose is to ensure that d_{jm} does not become too large if the value of a given predictor is close to 0. The second step in NSC involves determining the degree to which the individual predictors' group means should be shrunk toward the overall mean across groups. In order to do so, the value of the threshold parameter, Δ must be made. This is typically done using jackknife cross-validation in which a potential Δ is used and predictions are made for each jackknife sample. The threshold value that yields the most accurate cross-validation predictions is the one to be used in the final NSC algorithm. Shrinkage occurs by adjusting d_{jm} as in (9):

$$d'_{jm} = \text{sign}(d_{jm})(|d_{jm}|) - \Delta \quad (9)$$

Finally, (8) is solved for \bar{x}_{jm} and shrunken versions of the predictor means are calculated as in (10).

$$\bar{x}'_{jm} = \bar{x}_m + m_j(s_m + s_0)d'_{jm} \quad (10)$$

The shrunken centroids obtained in (10) are then used in (7). An important point to note here is that if for a given predictor, the shrinkage takes its centroid value down to (or past) 0, the centroid is assigned the value of 0. As an example, if

a predictor centroid is 1 and the amount of shrinkage determined using (9) and (10) is -2, then the shrunk centroid value would be 0, because 1-2 takes the value down to (and past) 0. NSC is known to have two advantages when used with high dimensional data. First, it reduces the impact of predictors with high variances, thus also reducing the amount of noise in the predictions themselves. Second, it creates a *de facto* predictor selection algorithm by removing the impact of variables that contribute relatively less to group separation (Tibshirani, Hastie, Narasimhan, & Chu, 2002).

Random Forest

The final method of classification to be considered in this study is the Random Forest (RF) of Brieman (2001), which is based upon the classification and regression tree (CART) recursive partitioning algorithm that Breiman, Friedman, Olshen, & Stone (1984) described. For CART with a categorical outcome variable, predictors are used to partition members of the sample in ever more heterogeneous groups, with respect to the outcome. The partitioning continues until a predetermined stopping rule has been reached such that no further divisions of the sample will yield appreciable gains in prediction accuracy.

A problem with CART is that it has a tendency to overfit the training data, making the resultant prediction algorithm less generalizable to the general population. However, it is also true that CART solutions are unbiased so that if they are averaged across a great many samples from the population, the results should provide very accurate prediction heuristics (Dietterich, 2000; Bauer & Kohavi, 1999). Brieman used this unbiasedness property in developing RF, which consists of an ensemble of CART results applied to a sample, and then averaged to create a single prediction algorithm. RF works by randomly selecting B subsamples of the original sample, either with replacement and therefore being of size n , or only a portion of the total sample without replacement so that the subsample is less than n . Those individuals not included in a subsample for a given tree are referred to as the out of bag sample. In addition, a subsample of the predictors is also randomly selected, and used with CART to create a prediction tree for the subsample of individuals. This process is completed a large number of times (e.g. 1000), and the resulting trees are saved after each analysis. Each tree is then applied to members of the training sample, or to new individuals, and a predicted outcome (e.g. classification) is obtained. These results are then averaged across the B trees for each individual in order to obtain a RF predicted value. The diversity of solutions introduced through the large number of trees based on subsamples of both

individuals and predictors results in a final solution that is more generalizable than any individual CART model.

Variable importance in prediction is determined through permutation tests (Nicodemus, Malley, Strobl, & Ziegler, 2010). For RF, the permutation importance of an individual predictor variable is calculated by comparing the number of correct predictions made by the actual data (i.e. the predictor ordered as it appears in the original dataset) with the number of correct predictions made when the variable has been permuted (i.e. randomly shuffled), averaged across all trees in the ensemble. The classification accuracy rate across trees for the original variable with no permutation is then compared with that of the mean accuracy rate for the permuted trees. If the difference in prediction accuracy is large, and presumably in favor of the tree based on the original data, we would conclude that the variable is important in accurately predicting group membership. On the other hand, if the difference in classification accuracy between the actual and permuted values is very small, then we would conclude that the variable does not contribute much more to determining group membership than if it were random and thus totally unrelated to the outcome. More formally, importance for variable x_m for a single tree (t) is calculated as:

$$VI_t(x_m) = \frac{\sum I(y_i = \hat{y}_{io})}{|\bar{B}|} - \frac{\sum I(y_i = \hat{y}_{ip})}{|\bar{B}|} \quad (11)$$

where

\hat{y}_{io} = Predicted class for observed data

\hat{y}_{ip} = Predicted class for permuted data

\bar{B} = out-of-bag sample

If variable x_m is not included in the tree, then $VI=0$. In order to obtain the overall variable importance measure for the RF, we then calculate

$$VI(X_m) = \frac{\sum_{t=1}^T VI_t(x_m)}{T} \quad (12)$$

where T is the total number of trees in the ensemble.

Methods

The research questions outlined above were addressed through the use of a Monte Carlo simulation study carried out with in the R software system, version 2.15.1 ([R Foundation for Statistical Computing, 2011](#)). The variables that were manipulated in the simulation study were selected in order to mirror conditions that researchers, particularly in the social sciences, might see when faced with a high dimensional dataset. For all conditions, data were simulated for two groups and unless otherwise noted the data were from a standard multivariate normal distribution.

Manipulated variables

Method A total of 6 methods were examined in this study, including LDA, RF, PLDA, SDA, NSC, and RDA. For methods relying on the setting of tuning parameters for optimal performance, the jackknife cross-validation methods described above were incorporated into the simulation code.

Sample size Sample size conditions included in the study were 10, 20, 30, 40, and 50. In all cases group sizes were held equal.

Number of predictors The number of predictors simulated in this study were 14, 28, and 50. Taken together with the sample size conditions discussed above, the ratios of P to N ranged from 5/1 to just over 1/4. While these conditions would not be considered terribly high dimensional in genetics, or another science where extreme high dimensionality is common, they do represent relatively high dimensional data in the context of psychology, education, and other social sciences, in which researchers typically strive to have many more subjects than variables.

Group mean separation The separation between group means was quantified in terms of Cohen's d effect size. For all predictors the groups' means differed by the same amount, either 0.2, 0.5, or 0.8. Thus, for example, in the $P=50$ group mean difference 0.5 case, all 50 variables were simulated to differ by 0.5 between the two groups. Group 1 was simulated to have means of 0 and standard deviations of 1 across conditions, and group 2 was simulated with means of 0.2, 0.5, or 0.8 and standard deviations of 1 for all predictors, depending on the group mean separation condition.

Correlation among predictors The predictors were simulated to have correlations among one another of 0, 0.5, or 0.8. These values were selected in order to assess performance of the methods in two relatively extreme cases (no correlation, very high correlation), and when the correlation was in the middle.

Distribution of the predictors In order to investigate the performance of the methods when data were normal and when they were not, two distribution conditions were simulated: multivariate normal and skewed with skewness of 2.5. Given the reliance of some of the methods on the assumption of normality, in particular LDA, it was of some interest to ascertain the impact that violating the assumption would have on performance of the methods.

Simulation outcomes Two outcome variables were examined in this study. First, the overall misclassification rate for a cross-validated sample drawn from the identical distribution as the training sample for a given combination of simulation conditions was recorded. This rate simply represents the proportion of cases that were incorrectly classified by each method. The second outcome variable of interest was the proportion of predictor variables that were correctly identified as being associated with group separation. As noted above, all predictors were simulated to differ between the groups, so in the population this proportion was 1 for every simulation condition. Therefore, this outcome variable reflects the proportion of predictors that each method correctly found to contribute to group differences. For LDA, a variable was considered to contribute to the classification solution if the absolute value of its structure value was 0.32 or greater (Tabachnick & Fidell, 2013). With respect to RF, variables were considered to be important if the permutation test statistic described above was statistically significant at $\alpha=0.05$. With regard to RDA, SDA, and PLDA variable importance was determined through the standardized discriminant weights. Based on findings in Cao, Boitard, and Besse (2011), variables were considered to be important predictors if these standardized values were greater than or equal to 0.1. Finally, with respect to NSC, a predictor was considered to contribute to the prediction if its weights were not shrunk to 0, again in keeping with recommendations in the literature (Christin, Hoefsloot, Smilde, Hoekman, Suits, Bischoff, & Horvatovich, 2013).

All simulation conditions were completely crossed with one another for a total of 324 different simulations. For each of these simulations, 1000 replications were generated and analyzed. In order to ascertain which main effects and interactions of the manipulated conditions contributed significantly to the outcome variables, repeated measures analysis of variance (ANOVA) models were used. For each

A COMPARISON OF METHODS FOR GROUP PREDICTION

combination of simulation conditions, the outcome variables were calculated for each of the methods studied here, for each replication. These outcomes were then averaged across the 1000 replications in order to create individual values for the two outcomes of interest. These values then served as the dependent variables in two separate ANOVA models (one for misclassification and one for proportion of predictors correctly identified). The within subjects variable was method of classification, and the between subjects variables were the other manipulated factors described above. In addition to the statistical significance of the main effects and interactions of these factors, the η^2 effect size was also used to identify model effects worthy of post hoc investigation. Main effects and interactions that were statistically significant, and which had η^2 of 0.1 or greater were considered “important”, because they were associated with at least 10% of the variance in the outcome variable.

Results

Classification accuracy

The ANOVA used to determine which of the manipulated factors or their interactions were related to overall classification accuracy. The interaction of method (M) by sample size (N) by correlation (C) was significantly associated with classification accuracy ($F_{40,952} = 3.621, p < 0.001, \eta^2 = 0.132$), as was the interaction of M by number of predictors (P) ($F_{8,472} = 11.937, p < 0.001, \eta^2 = 0.168$), and the interaction M by mean difference (D) ($F_{68,952} = 10.514, p < 0.001, \eta^2 = 0.429$). The overall misclassification rates by method, sample size, and correlation among the predictors appear in Table 1.

HOLMES FINCH

Table 1. Misclassification rates by Method, Sample Size (N), and Correlation among the predictor variables (C)

N	C	LDA	RF	PLDA	SDA	NSC	RDA
10	0	0.34	0.09	0.24	0.24	0.21	0.21
	0.5	0.39	0.13	0.23	0.23	0.35	0.37
	0.8	0.42	0.16	0.20	0.21	0.40	0.41
20	0	0.29	0.10	0.19	0.24	0.19	0.19
	0.5	0.37	0.12	0.18	0.24	0.35	0.35
	0.8	0.42	0.09	0.18	0.16	0.40	0.40
30	0	0.29	0.04	0.20	0.31	0.20	0.20
	0.5	0.37	0.07	0.17	0.13	0.36	0.36
	0.8	0.41	0.07	0.18	0.14	0.39	0.40
40	0	0.26	0.05	0.18	0.35	0.18	0.18
	0.5	0.36	0.08	0.18	0.17	0.35	0.35
	0.8	0.40	0.09	0.19	0.19	0.39	0.39
50	0	0.24	0.08	0.18	0.36	0.17	0.17
	0.5	0.37	0.10	0.16	0.21	0.36	0.36
	0.8	0.42	0.10	0.20	0.23	0.41	0.41
200	0	0.19	0.07	0.16	0.33	0.16	0.16
	0.5	0.36	0.08	0.18	0.30	0.36	0.36
	0.8	0.39	0.09	0.20	0.33	0.39	0.39

The results in Table 1 show that RF uniformly had the lowest misclassification rates of the methods studied here, across both sample size and correlation among the predictor variables. The highest misclassification rates belonged to LDA, particularly for the combination of N less than 40, and C of 0.5 or 0.8. For the combination of N less than 50 and C of 0.5 or 0.8, SDA had among the lowest misclassification rates, after RF, though when the predictors were uncorrelated, these rates were among the highest, particularly for larger sample sizes. Finally, PLDA did not exhibit increases in misclassification rates with increasing sample sizes, unlike SDA, and it generally had lower misclassification rates for C of 0.5 or 0.8 than any other method except for RF, and SDA with N less than 50. In short, PLDA generally maintained consistent misclassification rates at or just under 0.2 for the conditions simulated here.

A COMPARISON OF METHODS FOR GROUP PREDICTION

Table 2: Overall Misclassification Rates by Method and number of Predictors (P)

P	LDA	RF	PLDA	SDA	NSC	RDA
14	0.37	0.14	0.31	0.22	0.34	0.33
28	0.34	0.09	0.23	0.21	0.32	0.31
50	0.33	0.10	0.15	0.20	0.31	0.30

Table 2 includes the misclassification rates for M by P . Each of the approaches exhibited lower misclassification rates in the presence of more predictors. This effect was muted, however, for all of the methods except PLDA. In the latter case, the decrease in the proportion of misclassified cases was 0.16 from 14 to 50 predictors, whereas for the other methods, the decline in misclassification was never more than 0.04. In other words, the number of predictors included in the analysis had a much greater impact on the performance of PLDA than it did on any of the other methods studied here. Finally, Table 3 includes the overall misclassification rates for M by D . Across all methods, misclassification rates declined as differences in group means increased. This decline was particularly notable for PLDA, which produced a difference in misclassification of 0.29 between $D=0.2$ and $D=0.8$. Similarly, LDA, NSC, and RDA also evinced declines in misclassification of more than 0.2 between the smallest and largest group separation conditions. On the other hand, both RF and SDA displayed much smaller such declines, though for these methods as well, the rates declined with increasing group separation.

Table 3: Overall Misclassification Rates by Method and Difference in Group Means (D)

D	LDA	RF	PLDA	SDA	NSC	RDA
0.2	0.47	0.15	0.45	0.26	0.45	0.43
0.5	0.35	0.11	0.19	0.21	0.31	0.30
0.8	0.23	0.07	0.16	0.17	0.21	0.21

Correct Identification of Predictors Contributing to Group Separation

As with the misclassification rates, ANOVA used to determine which of the manipulated factors or their interactions were related to the proportion of predictors correctly identified as being associated with group separation. The interaction of M by P was significantly associated with the proportion of predictors correctly

identified as related to group differences ($F_{40,392} = 2.818$, $p < 0.001$, $\eta^2 = 0.223$). In addition, the interaction of M by D ($F_{68,952} = 10.514$, $p < 0.001$, $\eta^2 = 0.457$), and M by predictor distribution (PD) ($F_{4,94} = 17.556$, $p < 0.001$, $\eta^2 = 0.428$) were also significantly related to the proportion of predictors identified as important.

Table 4: Proportion of Predictors Associated with Group Differences Correctly Identified by Method and Number of Predictors (P)

P	LDA	RF	PLDA	SDA	NSC	RDA
14	0.19	0.53	0.18	0.53	0.85	0.83
28	0.003	0.41	0.08	0.38	0.66	0.73
50	0.001	0.08	0.05	0.24	0.57	0.55

Table 4 includes the proportion of the number of predictors by M and P . LDA consistently displayed among the lowest, if not the lowest proportion of predictors correctly identified. The next lowest rates belonged to PLDA, which performed similarly to LDA with $P=14$, and somewhat better for $P=28$ and 50. RF and SDA had comparable predictor identification rates for $P=14$ and 28, but the performance of RF fell more dramatically for $P=50$ than was the case for SDA. The best performers in terms of correctly identifying predictor variables associated with the group differences were NSC and RDA, each of which had proportions that were 0.2 or higher than their nearest competitors. For example, when $P=14$, both methods accurately identified over 80% of the predictors as being associated with group separation. This value dropped to 57% and 55%, respectively, when $P=50$, which represented more accurate performance than any of the other methods, even at their best, when $P=14$.

The proportion of predictors correctly identified by the method (M) and group mean separation (D) appears in Table 5.

A COMPARISON OF METHODS FOR GROUP PREDICTION

Table 5: Proportion of Predictors Correctly Identified by Method and Difference in Group Means (D)

D	LDA	RF	PLDA	SDA	NSC	RDA
0.2	0.47	0.15	0.45	0.26	0.45	0.43
0.5	0.35	0.11	0.19	0.21	0.31	0.30
0.8	0.23	0.07	0.16	0.17	0.21	0.21

As was evident in Table 4, across methods LDA had the lowest correct proportion of predictors, except for $D=0.8$, in which case PLDA had the lowest proportion. Indeed, the ability of PLDA to correctly identify the number of predictors associated with group membership did not seem to be associated with group separation, as its rate stayed large constant. RF and SDA had similar rates to one another for $D=0.2$ and $D=0.8$, but SDA performed somewhat better when $D=0.5$. Neither of these methods performed as well as RDA or NSC, however. RDA had the highest proportion of predictors correctly identified for both $D=0.2$ and 0.5 , and was slightly lower than NSC for $D=0.8$. Furthermore, the rates for RDA were largely unaffected by the degree of group separation, making it almost as accurate for low mean differences as for high ones. On the other hand, the performance of NSC was much more strongly influenced by D , as is evidenced by the change in the proportion of predictors from 0.2 to 0.8 .

Table 6 includes the proportion of predictors by PD.

Table 6: Proportion of Features Correctly Identified as Important by Method and Predictor Distribution

PD	LDA	RF	PLDA	SDA	NSC	RDA
Normal	0.13	0.38	0.20	0.43	0.66	0.98
S2.5	0	0.30	0.07	0.33	0.65	0.42

Several of the methods were deleteriously impacted by the presence of skewness in the distribution of predictors, in particular RDA, which was nearly perfect in identifying the correct number of important predictor variables when the data were normal, but did so less than half the time for skewed data. Similarly, LDA, RF, PLDA, and SDA all had proportions of predictor rates for the S2.5 condition 0.08 or more lower than was the case with normal data. On the other hand, the performance of NSC in terms of correctly identifying the number of predictors was virtually unaffected by predictor distribution.

Discussion

The goal of this simulation study was to compare several methods for supervised group classification in the presence of high dimensional data. Prior work in this area has tended to focus on a small number of such methods using applied examples with single datasets, or small simulation studies with relatively few manipulated conditions. The goal of this study was to expand upon these earlier efforts in several ways. First, by utilizing a larger set of simulated conditions than has been seen previously, we were able to test the various methods across a wider array of scenarios. In addition, we included a number of methods in this study that had not been previously compared with one another, including RF, which has never been systematically studied in the high dimensional case. Finally, this study examined the performance of the methods both in terms of their ability to correctly classify individuals into groups, and in terms of their use of salient predictors.

As described above, the results of this simulation study clearly support the use of RF if the primary goal of the researcher is to correctly classify individuals into their appropriate groups. No other method was nearly as effective in this regard, across all conditions simulated here. Conversely, standard LDA was the worst performer in terms of prediction accuracy, across virtually all conditions simulated here. The other approaches, each of which relied on some type of regularization or penalty function, produced misclassification rates between these two methods. In examining why RF might have performed so much better than the alternatives, we might consider its very nature as a recursive partitioning algorithm. As noted above, a problem with many prediction models in the high dimensional case is that the covariance matrices used to obtain model coefficients are ill behaved and sometimes singular. The regularization methods studied here (e.g. RDA, PLDA, SDA, NSC) each attempts to deal with this problem by reducing the number of predictors that are used in the prediction. However, in doing so, they also reduce the number of variables that contribute to group prediction, including those that might be salient. RF, on the other hand, does not use the covariance matrix at all, and thus does not face the problem of poor estimation of model coefficients faced by LDA, and reduction in the number of variables used in prediction that is a part of the regularized approaches. RF simply divides the sample based on the available data, selecting the best predictors at each step of the tree building process. Furthermore, because it relies on a large number of such trees, each of which is based upon a subset of the predictors and members of the sample, it should be more generalizable to the population than perhaps are some of the other methods. And indeed, we found this to be the case in the current study.

A COMPARISON OF METHODS FOR GROUP PREDICTION

While RF provided the most accurate predictions of group membership, it was not particularly effective at identifying the number of salient predictors of group separation. The permutation test used to do so is still fairly new and untried, and so while it has been shown to work reasonably well with larger samples (Nicodemus, Malley, Strobl, & Ziegler, 2010), there has been little work done with small samples, regardless of the number of predictors. Given that significance for a given predictor variable is determined by comparing classification accuracy using it in its natural state, and when it is randomly ordered, it is possible that with small samples and many predictors there is simply little difference in accuracy associated with any one variable. On the other hand, both NSC and RDA were much more accurate in terms of identifying the number of predictors associated with group separation. In considering which of these methods might be optimal if a researcher's goal is to identify variables associated with group separation, the results of this study would suggest that the decision should be based upon the nature of data being used. For example, if the researchers are unsure as to how different the predictor group means are, or if it is known that differences for some of them are relatively small, and the data are normally distributed, then RDA might be the best choice. Its ability to correctly identify the number of salient predictors was optimal when the data were multivariate normal, and it seemed largely uninfluenced by the degree of mean separation. In particular, it was the most effective approach when the effect size separating the groups was small. On the other hand, if the researcher knows that the data are not normally distributed, NSC might be the best approach to use because it was the least affected by the skewness simulated here. RDA performed relatively poorly in the presence of skewed data.

Recommendations and directions for future research

The results of this study suggest some recommendations for practice for researchers faced with high dimensional data. First, if the primary goal is to develop some type of prediction algorithm to be used with future cases, then RF seems to be the best choice. It provided much more accurate predictions than any of the other methods, regardless of the nature of the data. On the other hand, if the researcher is most interested in trying to identify which variables are most associated with group separation, then NSC or RDA may be better choices than RF. In particular, if the data are normally distributed, RDA would be recommended, whereas if the data are skewed then NSC is likely the optimal choice. In all cases, LDA is not recommended when the number of predictors approaches, or is larger than the sample size.

The current study represents an extension of prior work in this area in terms of the number of high dimensional prediction methods examined, and the number of conditions simulated. However, it also has limitations that future research should seek to address. First of all, only two groups were simulated here. Future studies in this area need to compare the performance of these methods with three or more groups. In addition, all of the variables were simulated to be related to group separation. However, in reality researchers are often faced with a situation in which only some of the variables are related to group differences. Therefore, future simulation studies should include some predictors that are not different between the groups. Finally, given the clear impact of predictor distribution on the accuracy of some methods, future studies should expand upon the nature of nonnormal data, including some categorical variables.

References

- Bauer, E. & Kohavi, R. (1999). An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants. *Machine Learning*, 36, 105-139.
- Breiman, L. (2001). Random Forests. *Machine Learning* 45(1), 5–32.
- Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). *Classification and Regression Trees*. New York: Wadsworth.
- Bühlmann, P. & van de Geer, S. (2011). *Statistics for High-Dimensional Data*. New York: Springer.
- Cao, K-A. L., Boitard, S., & Besse, P. (2011). Sparse PLS Discriminant Analysis: Biologically Relevant Feature Selection and Graphical Displays for Multiclass Problems. *BMC Bioinformatics*, 22, 253-263.
- Christin, C., Hoefsloot, H.C., Smilde, A.K., Hoekman, B., Suits, F., Bischoff, R., & Horvatovich, P. (2013). A Critical Assessment of Feature Selection Methods for Biomarker Discovery in Clinical Proteomics. *Molecular Cell Proteomics*, 12(1), 263-276.
- Clemmensen, L., Hastie, T., Witten, D., & Ersbøl, B. (2011). Sparse Discriminant Analysis. *Technometrics*, 53(4), 406-413.
- Dietterich, T. G. (2000). An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization. *Machine Learning*, 40, 139-157.

A COMPARISON OF METHODS FOR GROUP PREDICTION

- Guo, Y., Hastie, T. & Tibshirani, R. (2007). Regularized linear discriminant analysis and its application in microarrays. *Biostatistics*, 8, 86–100.
- Hastie, T., Buja, A., & Tibshirani, R. (1995). Penalized Discriminant Analysis. *Annals of Statistics*, 23, 73-102.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. New York: Springer.
- Huberty, C. J. & Olejnik, S. (2006). *Applied MANOVA and Discriminant Analysis*. Hoboken, NJ: John Wiley & sons, Inc.
- Kriegel, H-P, Kröger, P., & Zimek, A. (2009). Clustering high-dimensional data: A Survey on Subspace Clustering, Pattern-Based Clustering and Correlation Clustering. *ACM Transactions on Knowledge Discovery from Data*, 3(1), 1-58.
- Nicodemus, K. K., Malley, J. D., Strobl, C., & Ziegler, A. (2010). The Behavior of Random Forest Permutation Based Variable Importance Measures under Predictor Correlation. *Bioinformatics*, 11, 110-122.
- Palmer, G. A. (2006). Neuropsychological Profiles of Persons with Mental Retardation and Dementia. *Research in Developmental Disabilities: A multidisciplinary Journal*, 27(3), 299-308.
- R Foundation for Statistical Computing. (2011). R Software, V. 2.15.1. Vienna, Austria.
- Sanden, J. (2004). Math/FCS Class Boosts Test Scores. *Journal of Family and Consumer Sciences*, 96(1), 18-19.
- Sanden, C., Befus, C. & Zhang, J. Z. (2008). Clustering-Based Genre Prediction on Music Data. Paper presented at the annual meeting of the Association for Computing Machinery, Montreal, QC, May.
- Siklos, S. & Kerns, K. A. (2007). Assessing the Diagnostic Experiences of a Small Sample of Parents of Children with Autism Spectrum Disorders. *Research in Developmental Disabilities: A multidisciplinary Journal*, 28(1), 9-22.
- Tabachnick, B. G. & Fidell, L. S. (2013). *Using Multivariate Statistics*. Boston, MA: Pearson.
- Tibshirani, R., Hastie, T., Narasimhan, B. & Chu, G. (2002). Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proceedings of the National Academy of Science, USA*, 99, 6567-6572.
- Tibshirani, R., Saunders, M., Rosset, S., Zhu, J. & Knight, K. (2005). Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society, B*, 67, 91-108.

Witten, D. M. & Tibshirani, R. (2011). Penalized Classification using Fisher's Linear Discriminant. *Journal of the Royal Statistical Society, B*, 73, 753-772.

Xu, B., Huang, J. Z., Williams, G. J., Wang, Q., Ye, Y. (2012). Classifying Very High-Dimensional Data with Random Forests Built from Small Subspaces. *International Journal of Data Warehousing and Mining*, 8(2), 1-20.

Zhang, H., Yu, C-Y, Singer, B., & Xiong, M. (2001). Recursive Partitioning for Tumor Classification with Gene Expression Microarray Data. *Proceedings of the National Academy of Sciences*, 98(12), 6730-6735.

Zhang, Z., Dai, G., & Jordan, M. I. (2010). Regularized Discriminant Analysis, Ridge Regression, and Beyond. *Journal of Machine Learning Research*, 11, 2199-2228.

Zou, H. & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, B*, 67, 301–320.

Retained-Components Factor Transformation: Factor Loadings and Factor Score Predictors in the Column Space of Retained Components

André Beauducel
University of Bonn
Bonn, Germany

Frank Spohn
University of Hamburg
Hamburg, Germany

Factor loadings optimally account for the non-diagonal elements of the covariance matrix of observed variables. Principal component analysis leads to components accounting for a maximum of the variance of the observed variables. Retained-components factor transformation is proposed in order to combine the advantages of factor analysis and principal component analysis.

Keywords: Factor analysis, principal component analysis, exploratory factor analysis.

Introduction

Common factor analysis (FA) is regularly used in order to identify latent constructs accounting for the covariance of observed variables whereas principal components analysis (PCA) is primarily used in order to explain as much of the variance as possible with a minimum of components (Conway & Huffcutt, 2003; Fabrigar, Wegener, MacCallum, Strahan, 1999; Preacher & MacCallum, 2003). There is a broad literature referring to similarities and differences between FA and PCA (Bentler & De Leeuw, 2011; Ogasawara, 2003; Harris, 2001; Velicer & Jackson, 1990; Unkel & Trendafilov, 2010). It is also known that both methods can produce identical or extremely similar results under specific conditions (Schneeweiss, 1997; Schneeweiss & Mathes, 1995) and that PCA is often used as a substitute for FA (Sato, 1990). Nevertheless, an important difference between FA and PCA is that communalities or unique error variances of the variables are not estimated in PCA whereas they are

Dr. Beauducel is Professor for Psychological Methods, Psychological Assessment, and Evaluation. Email Dr. Beauducel at beauducel@uni-bonn.de. Email Mr. Spohn at frank.spohn@uni-hamburg.de.

estimated in FA (Harman, 1976; Tabachnick & Fidell, 2007). The estimation of communalities avoids an inflation of loadings in FA whereas an inflation of loadings regularly occurs in PCA (Widaman, 1993; Snook & Gorsuch, 1989).

Another difference between PCA and FA is related to the scores resulting from these methods: The component scores are clearly determined in PCA whereas the factor scores are indeterminate in FA (Guttman, 1955; Lovie & Lovie, 1995; Grice, 2001). Moreover, the component scores account for a maximum of the variance of the observed variables so that they represent an optimal data reduction. The principal components represent best summarizers for the observed variables (ten Berge & Kiers, 1997), which might be relevant for psychological assessment. By contrast, in FA different factor score predictors with different advantages and disadvantages have been proposed (Beauducel & Rabe, 2009; Krijnen, 2006; ten Berge, Krijnen, Wansbeek, & Shapiro, 1999), however, there is no factor score predictor that is an optimal summarizer of the observed variables. In consequence, a method that combines an optimal estimation of the loading size (without inflation) with scores that represent an optimal data reduction is not available. Researchers have to decide: If they want to have an optimal representation of a latent construct and the corresponding loadings, they should opt for FA, if they want to get optimal summarizers of the observed variables, they should use PCA. In the present paper we start from the idea that a researcher wants to get both: An optimal (not inflated) loading matrix representing the common variance of latent constructs adequately and optimal summarizers of the observed variables. A method that combines the estimation of loading magnitude of FA with the optimal data reduction of PCA could be the projection of the factor loadings on the column space of the loadings of the components retained in PCA. The focus on the components retained for rotation and interpretation is necessary because typically the number of components retained in PCA is considerably smaller than the number of observed variables. Accordingly, the transformation resulting in factor loadings in the column space of the retained PCA loadings is called ‘retained-components factor transformation’ (RFT). First, some definitions are given, then RFT is introduced and RFT-factor scores summarizing the observed variables like principal component scores are presented. Finally, some properties of RFT are described and illustrated by means of a small simulation study and by means of an empirical example.

Methodology

The principal component model

According to Hotelling (1933) it is possible to decompose \mathbf{x} , the random vector of observations of order p by means of a linear combination of components $\boldsymbol{\zeta}$ with component loadings \mathbf{L} . The observations \mathbf{x} and the components $\boldsymbol{\zeta}$ are assumed to have an expectation zero ($E[\mathbf{x}] = 0$, $E[\boldsymbol{\zeta}] = 0$)

$$\mathbf{x} = \mathbf{L}\boldsymbol{\zeta} \quad (1)$$

This decomposition by means of components represents a population model. The principal component representation implies that $\mathbf{L}'\mathbf{L}$ is diagonal with elements ordered from large to small (ten Berge & Knol, 1985). Moreover, it is assumed that $\mathbf{L} \neq \mathbf{0}$. When the principal component model is applied to sample data, it will be reasonable to distinguish between a random vector $\boldsymbol{\zeta}_r$ of order q representing the intended and substantial variance and therefore the components retained for interpretation and a random vector $\boldsymbol{\zeta}_n$ of order $p - q$ representing the unintended or trivial variances and therefore the components not retained for interpretation (Hotelling, 1933). Accordingly, it is necessary to distinguish between \mathbf{M} , the $p \times q$ loading matrix of the retained components and \mathbf{N} the $p \times (p - q)$ loading matrix of the components not retained with $\mathbf{L} = [\mathbf{M} \mid \mathbf{N}]$. This yields

$$\mathbf{x} = \mathbf{M}\boldsymbol{\zeta}_r + \mathbf{N}\boldsymbol{\zeta}_n \quad (2)$$

with $\mathbf{M}'\mathbf{N} = \mathbf{0}$ and $\mathbf{N}'\mathbf{M} = \mathbf{0}$ following from $\mathbf{L}'\mathbf{L}$ being diagonal. The population covariance matrix of the observed variables can be decomposed as follows:

$$\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}' = \mathbf{M}\mathbf{M}' + \mathbf{N}\mathbf{N}' \quad (3)$$

The common factor model

The common factor model that is assumed to hold in the population is given by

$$\mathbf{x} = \boldsymbol{\Gamma}\boldsymbol{\xi} = [\boldsymbol{\Lambda} \mid \boldsymbol{\Psi}]\boldsymbol{\xi} = \boldsymbol{\Lambda}\boldsymbol{\xi}_c + \boldsymbol{\Psi}\boldsymbol{\xi}_u \quad (4)$$

where \mathbf{x} is the random vector of observations of order p , ξ is the random vector of factors consisting of q common factors ξ_c and p orthogonal unique factors ξ_u . Λ is the factor pattern matrix of order p by q , Ψ is the $p \times p$ diagonal unique loading matrix. It is assumed that Ψ contains only positive values and that $\Lambda \neq \mathbf{0}$. The factors ξ are assumed to have an expectation zero ($E[\xi] = \mathbf{0}$) and the standard deviation of ξ is one. Moreover, the expectation of the covariance of ξ_c with ξ_u is zero. The covariance matrix Σ can be decomposed into

$$\Sigma = \Lambda\Phi\Lambda' + \Psi^2 \quad (5)$$

where Φ represents the q by q factor correlation matrix.

Results

Retained-components factor transformation

In order to transform the retained component matrix \mathbf{M} to be as similar as possible to the factor loading matrix Λ the following transformation was used:

$$\mathbf{MT} = \Lambda \quad (6)$$

with the transformation matrix \mathbf{T} and the factor loading matrix Λ as a target matrix, much like in procrustes rotation (Hurley & Cattell, 1962). Solving Equation 6 for \mathbf{T} yields

$$\mathbf{T} = (\mathbf{M}'\mathbf{M})^{-1}\mathbf{M}'\Lambda \quad (7)$$

Entering \mathbf{T} into Equation 6 yields Λ^* , because the transformed component loadings will in most cases be similar, but not identical to the target matrix Λ . Accordingly, Λ^* contains the loadings resulting from retained-components factor transformation (RFT),

$$\mathbf{M}(\mathbf{M}'\mathbf{M})^{-1}\mathbf{M}'\Lambda = \Lambda^* \quad (8)$$

Equation 2 and Equation 4 both explain the variance of the observed variables \mathbf{x} , so that they can be equated. This yields

RETAINED-COMPONENTS FACTOR TRANSFORMATION

$$\mathbf{M}\zeta_r + \mathbf{N}\zeta_n = \Lambda\xi_c + \Psi\xi_u \quad (9)$$

Premultiplication of Equation 9 with $\mathbf{M}(\mathbf{M}'\mathbf{M})^{-1}\mathbf{M}'$ yields

$$\begin{aligned} & \mathbf{M}(\mathbf{M}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{M}\zeta_r + \mathbf{M}(\mathbf{M}'\mathbf{M})^{-1}\mathbf{M}'\mathbf{N}\zeta_n \\ &= \mathbf{M}(\mathbf{M}'\mathbf{M})^{-1}\mathbf{M}'\Lambda\xi_c + \mathbf{M}(\mathbf{M}'\mathbf{M})^{-1}\mathbf{M}'\Psi\xi_u \end{aligned} \quad (10)$$

Since $\mathbf{M}'\mathbf{N} = \mathbf{0}$ and according to Equation 8 it is possible to write

$$\mathbf{M}\zeta_r = \Lambda^*\xi_c + \Psi^*\xi_u \quad (11)$$

with $\Psi^* = \mathbf{M}(\mathbf{M}'\mathbf{M})^{-1}\mathbf{M}'\Psi$. Equation 11 gives a factorial representation of the retained components ζ_r . Thus, each retained component is decomposed into a projection from the common factors and from the unique factors. According to Equation 2 it is possible to write

$$\mathbf{x} = \Lambda^*\xi_c + \Psi^*\xi_u + \mathbf{N}\zeta_n \quad (12)$$

Thus, the RFT has two error terms: One term representing the unique error of the factorial decomposition of the retained-components and the other error term represents the residual PCA components (i.e., those components that are not retained for interpretation). The covariance matrix of observed variables can be computed from RFT by means of

$$\begin{aligned} \Sigma &= (\Lambda^*\xi_c + \Psi^*\xi_u + \mathbf{N}\zeta_n)(\Lambda^*\xi_c + \Psi^*\xi_u + \mathbf{N}\zeta_n)' \\ &= \Lambda^*\Phi\Lambda^{*'} + \Lambda^*\xi_c\zeta_n'\mathbf{N}' + \Psi^{*s} + \Psi^*\xi_u\zeta_n'\mathbf{N}' + \mathbf{N}\mathbf{N}' \end{aligned} \quad (13)$$

Postmultiplication of Equation 9 with ξ_c' , subsequent premultiplication with $(\mathbf{N}'\mathbf{N})^{-1}\mathbf{N}'$ and transposing yields

$$\xi_c\zeta_n' = \Phi\Lambda'\mathbf{N}(\mathbf{N}'\mathbf{N})^{-1} \quad (14)$$

Postmultiplication of Equation 9 with ξ_u , subsequent premultiplication with $(\mathbf{N}'\mathbf{N})^{-1}\mathbf{N}'$ and transposing yields

$$\xi_u \zeta_n' = \Psi \mathbf{N} (\mathbf{N}'\mathbf{N})^{-1} \quad (15)$$

Entering Equation 14 and 15 into Equation 13 and some transformation yields

$$\begin{aligned} \Sigma &= \Lambda^* \Phi \Lambda^{*'} + \Psi^{*2} + \mathbf{N}\mathbf{N}' + \mathbf{M}(\mathbf{M}'\mathbf{M})^{-1} \mathbf{M}'(\Lambda \Phi \Lambda' + \Psi^2) \mathbf{N} (\mathbf{N}'\mathbf{N})^{-1} \mathbf{N}' \\ &= \Lambda^* \Phi \Lambda^{*'} + \Psi^{*2} + \mathbf{N}\mathbf{N}' + \mathbf{M}(\mathbf{M}'\mathbf{M})^{-1} \mathbf{M}'(\mathbf{M}\mathbf{M}' + \mathbf{N}\mathbf{N}') \mathbf{N} (\mathbf{N}'\mathbf{N})^{-1} \mathbf{N}' \\ &= \Lambda^* \Phi \Lambda^{*'} + \Psi^{*2} + \mathbf{N}\mathbf{N}' + \mathbf{M}\mathbf{M}' \mathbf{N} (\mathbf{N}'\mathbf{N})^{-1} \mathbf{N}' \\ &= \Lambda^* \Phi \Lambda^{*'} + \Psi^{*2} + \mathbf{N}\mathbf{N}' \end{aligned} \quad (16)$$

since $\mathbf{M}'\mathbf{N} = \mathbf{0}$. Thus, the residual covariances that are represented by the loadings of the irrelevant components \mathbf{N} have no covariance with Λ^* of RFT, since $\mathbf{N}'\mathbf{M} = \mathbf{0}$ implies $\mathbf{N}'\mathbf{M}(\mathbf{M}'\mathbf{M})^{-1}\mathbf{M}'\Lambda = \mathbf{0}$. In contrast, the residual covariances represented by \mathbf{N} might be related to the FA-loadings, that is, $\mathbf{N}'\Lambda$ is not necessarily zero. One would therefore expect that advantages of RFT over conventional FA in terms of stability of parameters occur when the PCA residuals in $\mathbf{N}\mathbf{N}'$ primarily represent covariances due to sampling error. Moreover, RFT should help to avoid the overestimation of loadings as it occurs with PCA, because the PCA loadings are transformed in order to be as similar as possible to the factor loadings. Accordingly, a simulation study was performed in order to explore the quality of the sample RFT-loadings as estimators of population factor loadings.

Properties of the Retained-components factor score predictors

A main reason for proposing RFT was that it allows for factor score predictors that are optimal summarizers of the observed variables. This property holds for Harman's ideal-variable factor score predictor. The weights for Harman's (1976) ideal-variable factor score predictor based on RFT are given by

$$\mathbf{B}_H = \Lambda^* \left(\Lambda^{*'} \Lambda^* \right)^{-1'} \quad (17)$$

RETAINED-COMPONENTS FACTOR TRANSFORMATION

Moreover, according to Equation 2, the weights of the retained components are given by

$$\mathbf{B}_r = \mathbf{M}(\mathbf{M}'\mathbf{M})^{-1} \quad (18)$$

The relation between the component scores and Harman's ideal variables scores for RFT can be expressed in terms of correlations. Therefore, the correlations between weighted composites of the observed variables can be computed. Entering \mathbf{B}_H and \mathbf{B}_r into the formula for the correlations between weighted composites (Harris, 2001), yields

$$\mathbf{B}_H' \Sigma \mathbf{B}_r (\mathbf{B}_H' \Sigma \mathbf{B}_H \mathbf{B}_r' \Sigma \mathbf{B}_r)^{-0.5} = \mathbf{C}_{Hr} \quad (19)$$

where the main diagonal of \mathbf{C}_{Hr} contains the correlation matrix between Harman's factor score predictor based on RFT and the retained principal components. Entering Equation 17 and 18 into Equation 19 and some transformation yields

$$\mathbf{G}^{-1} (\mathbf{M}'\mathbf{M})^{-1} \mathbf{M}' \Sigma \mathbf{M} (\mathbf{M}'\mathbf{M})^{-1} \left(\mathbf{G}^{-1} (\mathbf{M}'\mathbf{M})^{-1} \mathbf{M}' \Sigma \mathbf{M} (\mathbf{M}'\mathbf{M})^{-1} \right)^{-1} = \mathbf{G}^{-1} \mathbf{G} = \mathbf{I} \quad (20)$$

with $\mathbf{G} = (\mathbf{M}'\mathbf{M})^{-1} \mathbf{M}' \Lambda$. Thus, the correlations between the component scores and Harman's ideal variables scores are all perfect for the RFT-solution. This implies that Harman's ideal variables scores of the RFT-solution are optimal summarizers of the observed variables as are the principal components.

Since Thurstone's (1935) least squares regression score predictor is often used and recommended (Krijnen, 2006), the relationship between the regression score predictor based on RFT and the principal component scores was explored. Since the principal component scores have the interesting property of being the optimal summarizers of the observed variables, they should be regarded as a criterion and the RFT regression factor scores as predictors. Thus, q multiple regressions and corresponding multiple correlations can be calculated for the q retained components. If the multiple correlations between the regression score predictors and the principal component scores as criterion is one, this indicates that the scores represent the same overall individual differences, even though they might be distributed differently on the factors and components. The weights for Thurstone's regression factor score predictor based on RFT are

$$\mathbf{B}_T = \Sigma^{-1} \Lambda^* \Phi = \Sigma^{-1} \mathbf{M}(\mathbf{M}'\mathbf{M})^{-1} \mathbf{M}'\Lambda\Phi \quad (21)$$

The corresponding regression weights for the prediction of the retained principal components from Thurstone's regression factor score predictor are

$$\mathbf{B} = (\Phi\Lambda^* \Sigma^{-1} \Lambda^* \Phi)^{-1} \Phi\Lambda^* \mathbf{M}(\mathbf{M}'\mathbf{M})^{-1} \quad (22)$$

The multiple correlation is calculated as

$$\begin{aligned} \mathbf{R}^2 &= \mathbf{B}'\Phi\Lambda^* \mathbf{M}(\mathbf{M}'\mathbf{M})^{-1} (\mathbf{M}'\mathbf{M})^{-1} \mathbf{M}'\Sigma\mathbf{M}(\mathbf{M}'\mathbf{M})^{-1} \\ &= (\mathbf{M}'\mathbf{M})^{-1} \mathbf{M}'\Lambda^* \Phi(\Phi\Lambda^* \Sigma^{-1} \Lambda^* \Phi)^{-1} \Phi\Lambda^* \mathbf{M}(\mathbf{M}'\mathbf{M})^{-1} \end{aligned} \quad (23)$$

Some transformation yields

$$\mathbf{R}^2 = (\mathbf{M}'\Sigma^{-1}\mathbf{M})^{-1} \quad (24)$$

A singular value decomposition of Σ yields $\Sigma = \mathbf{S}\mathbf{D}\mathbf{S}'$, with \mathbf{D} containing a diagonal matrix of eigenvalues in descending order. Accordingly it is possible to write

$$\mathbf{L}\mathbf{L}' = \mathbf{S}\mathbf{D}^{1/2}\mathbf{D}^{1/2}\mathbf{S}' = \mathbf{S} \quad (25)$$

From $\Sigma^{-1} = (\mathbf{S}\mathbf{D}^{-1}\mathbf{S}')'$ we get $\mathbf{L}\Sigma^{-1}\mathbf{L}' = \mathbf{S}\mathbf{D}^{1/2}\mathbf{D}^{-1}\mathbf{D}^{1/2}\mathbf{S}' = \mathbf{I}_{p \times p}$, which implies

$$\mathbf{M}\Sigma^{-1}\mathbf{M}' = \mathbf{S}\mathbf{D}^{1/2}\mathbf{D}^{-1}\mathbf{D}^{1/2}\mathbf{S}' = \mathbf{I}_{q \times q} \quad (26)$$

and, accordingly, $(\mathbf{M}'\Sigma^{-1}\mathbf{M})^{-1} = \mathbf{I}_{q \times q}$, which implies that all multiple correlations with the RFT regression factor scores as predictors and each principal component as criterion are one.

Simulation Study

The expectation that the RFT-loadings are more stable than the FA-loadings when the residual covariances represent sampling error was investigated by means of a small simulation study based on orthogonal and oblique three-factor models. For

RETAINED-COMPONENTS FACTOR TRANSFORMATION

the Models 1 to 4, the population FA-loadings were identical to the population RFT-loadings. Schneeweiss and Mathes (1995) have shown that factor loadings can be perfectly transformed into the retained component loadings when all unique factor loadings are equal. Whenever the retained component loadings can be perfectly transformed into the factor loadings, it follows from Equations 6 and 7 that the RFT-loadings will be identical to the factor loadings ($\Lambda^* = \Lambda$), because Equation 7 yields the transformation matrix for the transformation of the retained component loadings into factor loadings. The condition of equal uniqueness of all variables holds for population Models 1 and 2. Moreover, population Models 1 to 4 represent a perfect simple structure (independent clusters) where all non-salient loadings are zero and the salient loadings on each factor are identical even when there are different salient loadings on different factors for Models 3 and 4 (see Table 1). This implies that multiplication with a scalar will allow to transform each vector of factor loadings into the corresponding component loadings. Again, a perfect transformation of retained component loadings into factor loadings implies that the RFT-loadings and the factor loadings are identical.

Whereas Model 1 represents an orthogonal perfect simple structure with large salient loadings Model 2 represents an orthogonal perfect simple structure with moderate salient loadings. Model 3 represents an oblique perfect simple structure with large salient loadings and Model 4 represents an oblique perfect simple structure with moderate salient loadings (see Table 1).

BEAUDUCEL & SPOHN

Table 1: Population loadings for models with identical FA- and RFT-loadings

Model 1			Model 2		
.700	.000	.000	.500	.000	.000
.700	.000	.000	.500	.000	.000
.700	.000	.000	.500	.000	.000
.700	.000	.000	.500	.000	.000
.700	.000	.000	.500	.000	.000
.000	.700	.000	.000	.500	.000
.000	.700	.000	.000	.500	.000
.000	.700	.000	.000	.500	.000
.000	.700	.000	.000	.500	.000
.000	.700	.000	.000	.500	.000
.000	.000	.700	.000	.000	.500
.000	.000	.700	.000	.000	.500
.000	.000	.700	.000	.000	.500
.000	.000	.700	.000	.000	.500
.000	.000	.700	.000	.000	.500
Model 3			Model 4		
.714	.000	.000	.520	.000	.000
.714	.000	.000	.520	.000	.000
.714	.000	.000	.520	.000	.000
.714	.000	.000	.520	.000	.000
.714	.000	.000	.520	.000	.000
.000	.665	.000	.000	.472	.000
.000	.665	.000	.000	.472	.000
.000	.665	.000	.000	.472	.000
.000	.665	.000	.000	.472	.000
.000	.665	.000	.000	.472	.000
.000	.000	.616	.000	.000	.424
.000	.000	.616	.000	.000	.424
.000	.000	.616	.000	.000	.424
.000	.000	.616	.000	.000	.424
.000	.000	.616	.000	.000	.424
inter-factor correlations					
1.000			1.000	-.061	.363
-.032	1.000		-.061	1.000	-.475
.273	-.329	1.000	.363	-.475	1.000

RETAINED-COMPONENTS FACTOR TRANSFORMATION

Population FA-loadings and the corresponding RFT-loadings for models with unequal FA- and RFT-loadings ($\Lambda \neq \Lambda^*$) are given in Table 2 and 3. Models 5 and 6 are orthogonal, Model 5 has a simple structure with large salient loadings, and Model 6 has a simple structure with moderate salient loadings (see Table 2). Moreover, Model 7 has an oblique simple structure and high salient loadings whereas Model 8 represents an oblique simple structure with low to moderate salient loadings (see Table 3). The eight models with their corresponding population factor loading matrices presented in Tables 1, 2, and 3 were used in order to generate population correlation matrices according to Equation 5. It should be noted that even for those population models where the FA- and RFT-loadings were not equal, the means of the FA- and the RFT-loadings were generally similar (see Tables 2 and 3, bottom). The only exception was found for the first factor of Model 7, where the mean RFT-loading was a bit smaller than the mean factor loading. Overall, this demonstrates that the RFT-loadings are not inflated.

From each population 500 random normal samples with 50, 75, 150, 300, and 1000 cases were taken. Maximum likelihood factor analysis (MLFA), unweighted least squares factor analysis (ULFA), and PCA were performed for each sample correlation matrix. It should be noted that the relative size of the MLFA-loadings does not depend on the standard deviations of the observed variables, which means that MLFA is scale free (Lawley, 1940). On the other hand, PCA is not scale free so that the relative size of the PCA-loadings can be affected by different standard deviations of the observed variables. Since RFT is based on PCA, it is not recommended to calculate RFT for ML-factors when covariance matrices are analyzed. Therefore, the present simulation study was based on correlation matrices so that no effects of scaling on the loadings were expected. It was decided to include a correlation-based MLFA into the simulation study, because ML-estimation is rather common in the context of factor analysis.

BEAUDUCEL & SPOHN

Table 2: Population loading matrices of Model 5 and 6

FA-loadings			RFT-loadings		
Model 5					
	.600	.000	.000	.638	.000
	.650	.000	.000	.673	.000
	.700	.000	.000	.705	.000
	.750	.000	.000	.734	.000
	.800	.000	.000	.759	.000
	.000	.600	.000	.000	.638
	.000	.650	.000	.000	.673
	.000	.700	.000	.000	.705
	.000	.750	.000	.000	.734
	.000	.800	.000	.000	.759
	.000	.000	.600	.000	.638
	.000	.000	.650	.000	.673
	.000	.000	.700	.000	.705
	.000	.000	.750	.000	.734
	.000	.000	.800	.000	.759
M	.233	.233	.233	.234	.234
Model 6					
	.400	.000	.000	.437	.000
	.450	.000	.000	.474	.000
	.500	.000	.000	.507	.000
	.550	.000	.000	.535	.000
	.600	.000	.000	.559	.000
	.000	.400	.000	.000	.437
	.000	.450	.000	.000	.474
	.000	.500	.000	.000	.507
	.000	.550	.000	.000	.535
	.000	.600	.000	.000	.559
	.000	.000	.400	.000	.437
	.000	.000	.450	.000	.474
	.000	.000	.500	.000	.507
	.000	.000	.550	.000	.535
	.000	.000	.600	.000	.559
M	.167	.167	.167	.167	.167

Note. "M" denotes the column mean.

RETAINED-COMPONENTS FACTOR TRANSFORMATION

Table 3: Population loading matrices of Model 7 and 8

FA-loadings			RFT-loadings			
Model 7						
	.673	.000	.000	.654	.001	.005
	.694	.000	.000	.666	.000	.002
	.714	.000	.000	.677	.000	.000
	.734	.000	.000	.688	.001	.002
	.755	.000	.000	.698	.001	.005
	.000	.624	.000	.001	.636	.006
	.000	.644	.000	.000	.648	.003
	.000	.665	.000	.000	.659	.000
	.000	.686	.000	.001	.671	.003
	.000	.706	.000	.001	.681	.006
	.000	.000	.573	.005	.006	.586
	.000	.000	.594	.003	.003	.597
	.000	.000	.616	.000	.000	.609
	.000	.000	.638	.003	.004	.620
	.000	.000	.659	.006	.007	.629
M	.238	.222	.205	.227	.221	.205
inter-factor correlations						
	1.000			1.000		
	-.031	1.000		-.035	1.000	
	.271	-.327	1.000	.265	-.340	1.000
Model 8						
	.478	.000	.000	.496	.002	.007
	.498	.000	.000	.508	.001	.003
	.520	.000	.000	.521	.000	.000
	.540	.000	.000	.531	.001	.004
	.561	.000	.000	.541	.002	.007
	.000	.428	.000	.002	.446	.010
	.000	.450	.000	.001	.458	.005
	.000	.471	.000	.000	.468	.001
	.000	.492	.000	.002	.478	.006
	.000	.514	.000	.003	.487	.011
	.000	.000	.376	.009	.012	.397
	.000	.000	.399	.004	.006	.407
	.000	.000	.423	.001	.001	.417
	.000	.000	.446	.006	.008	.425
	.000	.000	.470	.011	.015	.433
M	.173	.157	.141	.176	.159	.142
inter-factor correlations						
	1.000			1.000		
	-.060	1.000		-.067	1.000	
	.359	-.468	1.000	.340	-.505	1.000

Note. "M" denotes the column mean.

Varimax-rotation was performed for the orthogonal models (Model 1, 2, 5, and 6) and Promax-rotation ($Kappa=4$) was performed for the oblique models (Model 3, 4, 7, and 8). Then, according to Equation 8 the RFT-loadings were computed from the unrotated sample PCA retained component loadings and the factor loadings. Varimax-rotation of the RFT-loadings was performed for the orthogonal models and Promax-rotation ($Kappa=4$) was performed for the oblique models.

Although the RFT constitutes a new model comprising aspects both from PCA and FA, researchers might want to use the RFT especially as a substitute for FA. Therefore, the root mean square (RMS) difference between the sample FA-loadings and the corresponding population FA-loadings was compared with the RMS difference between the sample RFT-loadings and the corresponding population FA-loadings (Figures 1 and 2). The RMS difference represents the overall difference between sample and population FA-loadings, but it does not indicate whether an over- or underestimation occurs. Therefore, the mean-difference between the mean sample loadings and the population FA-loadings was also calculated (see Table 4). The mean-difference is negative when the sample RFT-, FA-, or PCA-loadings underestimate the population FA-loadings and it is positive when the sample loadings overestimate the population FA-loadings.

Figure 1 contains the RMS differences between the sample MLFA-loadings, sample ULFA-loadings, sample ML-RFT-loadings, sample UL-RFT-loadings, sample PCA-loadings and the corresponding population FA-loadings for Models 1 to 4. RMS differences were equal or smaller for RFT-loadings based on ML-estimation than for MLFA-loadings. Moreover, RMS differences were equal or smaller for RFT-loadings based on UL-estimation than for ULFA-loadings. Thus, when the population FA-loadings and the population RFT-loadings are equal, the precision of the sample RFT-loadings as estimates of the population FA-loadings is at least as high as the precision of the FA-loadings. The mean-differences between sample MLFA-loadings, sample ULFA-loadings, sample RFT-loadings and the corresponding population FA-loadings were extremely small for Models 1 to 3 (see Table 4). They were a bit larger for Model 4, where a slight tendency for an underestimation of loadings was found for all methods. PCA-loadings have, in general, the largest RMS and, thus, the lowest precision as estimates of the FA-loadings, especially for sample sizes of 150 cases and above (see Figure 1) and the mean-differences were of a relevant size (see Table 4), indicating the known tendency of PCA-loadings to overestimate the population FA-loadings.

RETAINED-COMPONENTS FACTOR TRANSFORMATION

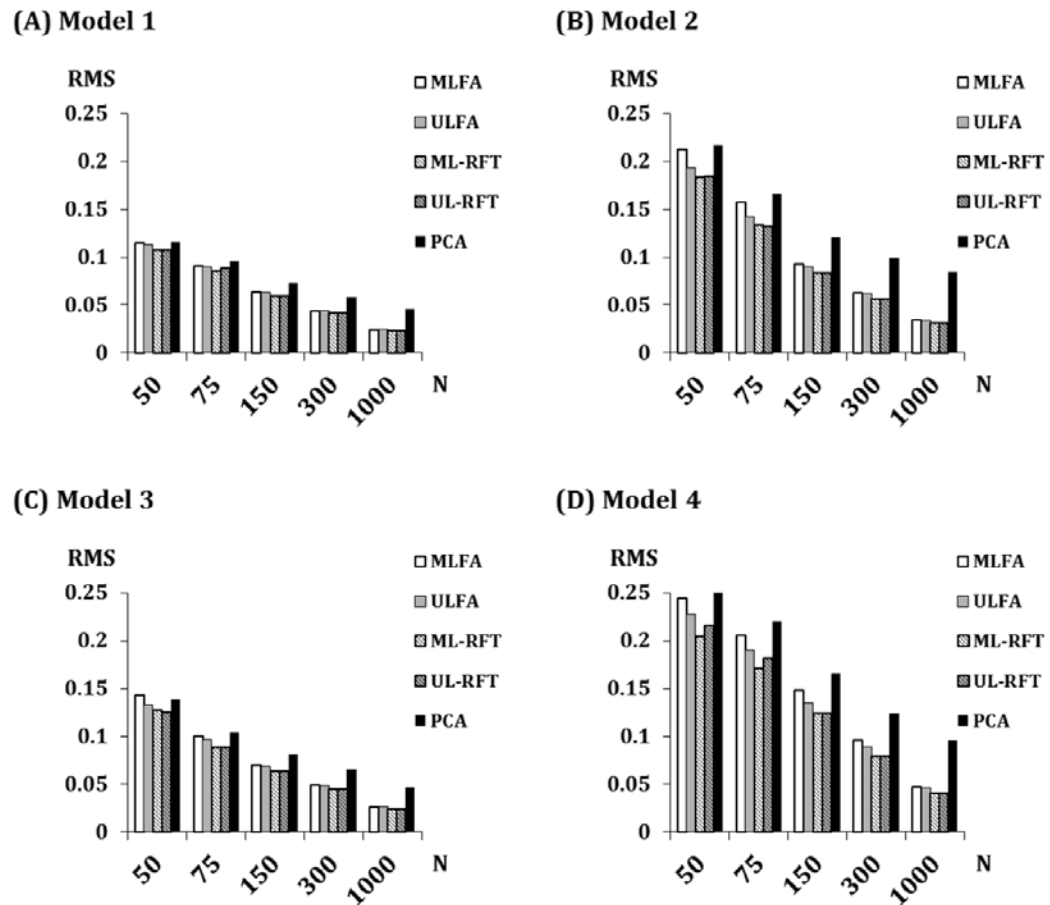


Figure 1. Root mean squared difference (RMS) between the sample MLFA-, ULFA-, ML-RFT-, UL-RFT-, PCA- loadings and the corresponding population factor loadings for Models 1 to 4.

BEAUDUCEL & SPOHN

Table 4. Mean-difference between sample loading estimates and population FA-loadings

Model	N	MLFA	ML-RFT	ULFA	UL-RFT	PCA
1	50	-.003	-.002	-.002	-.001	.019
	75	-.002	-.001	-.001	-.001	.021
	150	-.001	.000	-.001	.000	.022
	300	.000	.000	.000	.000	.023
	1000	.000	.000	.000	.000	.023
2	50	-.002	-.008	.000	.001	.032
	75	-.002	-.001	.001	.001	.036
	150	.000	.001	.001	.001	.041
	300	.000	.001	.000	.001	.043
	1000	.000	.000	.000	.000	.044
3	50	-.005	-.002	-.003	-.001	.021
	75	-.003	-.001	-.002	-.001	.023
	150	-.001	.000	-.001	.000	.025
	300	.000	.000	.000	.000	.026
	1000	.000	.000	.000	.000	.026
4	50	-.020	-.045	-.016	-.014	.013
	75	-.017	-.024	-.015	-.012	.020
	150	-.012	-.007	-.006	-.006	.032
	300	-.006	-.003	-.002	-.002	.041
	1000	-.002	-.002	-.001	-.001	.046
5	50	-.002	-.002	-.002	-.001	.019
	75	-.002	-.001	-.001	-.001	.020
	150	-.001	-.001	-.001	.000	.021
	300	.000	.000	.000	.000	.022
	1000	.000	.001	.000	.001	.023
6	50	-.001	-.018	.001	.002	.033
	75	.000	.001	.001	.002	.036
	150	.000	.001	.000	.002	.040
	300	.000	.001	.000	.001	.042
	1000	.000	.001	.000	.001	.043
7	50	-.021	-.009	-.010	-.008	.015
	75	-.012	-.009	-.010	-.008	.015
	150	-.009	-.007	-.008	-.007	.018
	300	-.006	-.006	-.006	-.006	.021
	1000	-.006	-.006	-.006	-.006	.021
8	50	-.017	-.044	-.015	-.014	.013
	75	-.018	-.017	-.016	-.015	.017
	150	-.013	-.009	-.011	-.008	.031
	300	-.007	-.004	-.006	-.003	.040
	1000	-.003	-.002	-.004	-.002	.046
M		-.005	-.006	-.004	-.003	.028

Note. "M" denotes the column mean.

RETAINED-COMPONENTS FACTOR TRANSFORMATION

The RMS differences between population FA-loadings and corresponding sample FA-loadings, sample RFT-loadings, and sample PCA-loadings were presented for Models 5 to 8 in Figure 2. Both for ML- and UL-estimation, the RMS differences were smaller for the RFT-loadings than for the FA-loadings. Although the population RFT-loadings were different from the population FA-loadings for Models 5 to 8, the sample RFT-loadings were at least as precise estimators of the population FA-loadings as the sample FA-loadings. The mean-differences between sample and population loadings were extremely small for MLFA-, ULFA-, ML-RFT-, and UL-RFT-loadings in Models 5 and 6. They tend to be a bit more negative for Model 7 and especially for Model 8 for samples comprising 50 and 75 cases (see Table 4). The overall mean-difference between UL-based RFT-loadings and population factor loadings was slightly smaller than the overall mean-difference for any other method (see Table 4, bottom). Again, the PCA-loadings had the lowest precision as estimates of the population FA-loadings both in terms of RMS (Figure 2) and in terms of the mean-differences, which indicate the overestimation of population FA-loadings by means of PCA (Table 4).

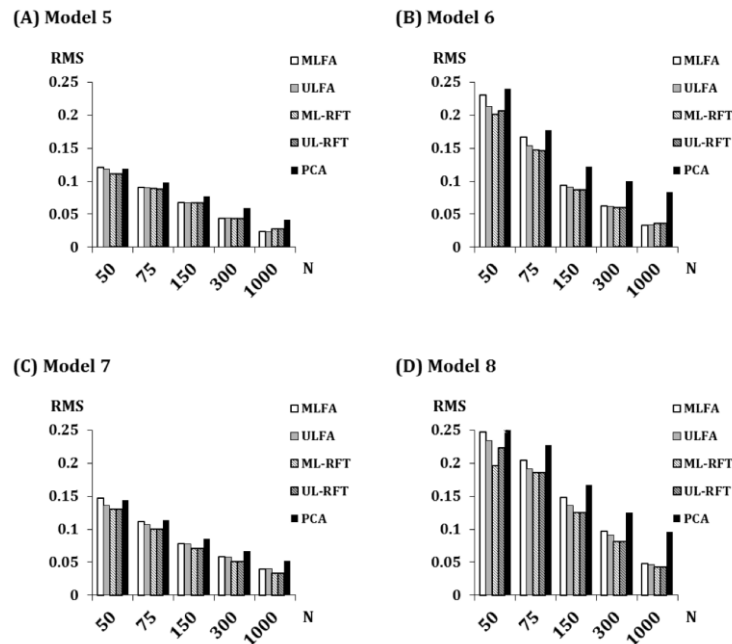


Figure 2. Root mean squared difference (RMS) between the sample MLFA-, ULFA-, ML-RFT-, UL-RFT-, PCA- loadings and the corresponding population factor loadings for Models 5 to 8.

The RMS differences between the inter-correlations of the population factors and the sample inter-correlations for MLFA, ULFA, the corresponding RFT, and the sample principal components were presented for the oblique population models (Model 3, 4, 7, and 8, see Figure 3). For Models 4, 7, and 8 and sample sizes below 150 cases the RMS differences were smaller for MLFA than for the RFT based on ML-estimation. Especially, when based on 50 cases, the RMS was large for the ML-based RFT for Models 4 and 8. However, this effect did not occur for the UL-based RFT. In contrast, when sample size was at least 150 cases the RMS was smaller for the ML-based RFT than for MLFA. For UL-based RFT the RMS tends to be equal or smaller than for ULFA. Overall, the mean-differences between sample inter-correlations and population factor inter-correlations indicate that the correlations tend to be underestimated (see Table 5). The effect of underestimation was most pronounced for PCA. Moreover, the underestimation of inter-factor correlations was less pronounced for RFT-solutions than for the FA-solutions with all methods (see Table 5).

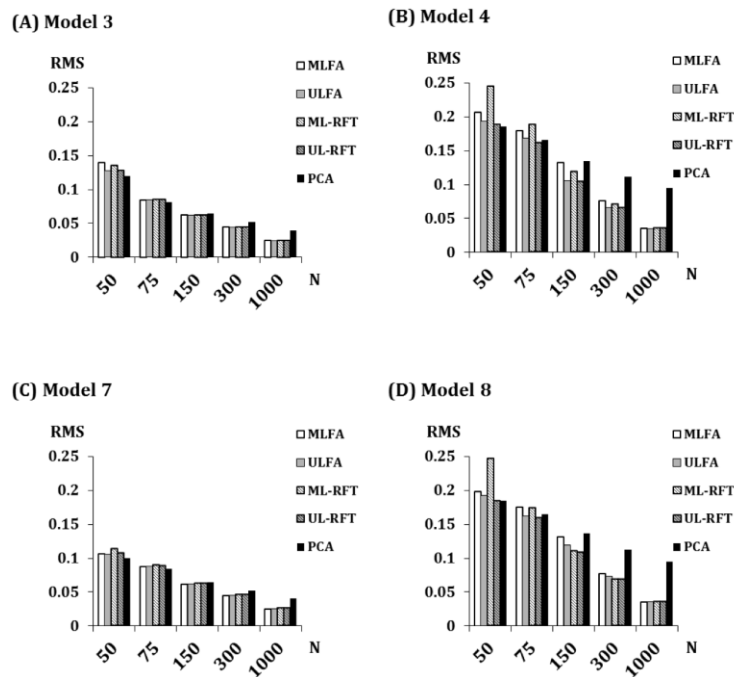


Figure 3. Root mean squared difference (RMS) between the inter-correlations for sample MLFA-, ULFA-, ML-RFT-, UL-RFT-, PCA and the corresponding population factor inter-correlations for the oblique models (Model 3, 4, 7, and 8).

RETAINED-COMPONENTS FACTOR TRANSFORMATION

Table 5. Mean-difference between sample factor and component inter-correlations and population inter-factor correlations for the oblique models (Model 3, 4, 7, and 8)

Model	N	MLFA	ML-RFT	ULFA	UL-RFT	PCA
3	50	-0.054	-0.039	-0.045	-0.041	-0.073
	75	-0.023	-0.017	-0.022	-0.019	-0.057
	150	-0.014	-0.011	-0.013	-0.011	-0.052
	300	-0.007	-0.006	-0.007	-0.006	-0.049
	1000	-0.003	-0.003	-0.003	-0.003	-0.047
4	50	-0.226	-0.167	-0.21	-0.194	-0.232
	75	-0.189	-0.112	-0.17	-0.151	-0.207
	150	-0.124	-0.076	-0.105	-0.082	-0.174
	300	-0.06	-0.038	-0.053	-0.041	-0.15
	1000	-0.017	-0.014	-0.017	-0.014	-0.135
7	50	-0.047	-0.034	-0.041	-0.038	-0.072
	75	-0.027	-0.019	-0.025	-0.021	-0.061
	150	-0.013	-0.01	-0.012	-0.01	-0.053
	300	-0.007	-0.005	-0.007	-0.005	-0.05
	1000	-0.002	-0.001	-0.002	-0.001	-0.047
8	50	-0.211	-0.151	-0.207	-0.189	-0.228
	75	-0.183	-0.138	-0.166	-0.15	-0.208
	150	-0.121	-0.079	-0.108	-0.087	-0.177
	300	-0.061	-0.04	-0.056	-0.042	-0.151
	1000	-0.018	-0.012	-0.017	-0.013	-0.135
M		-0.07	-0.049	-0.064	-0.056	-0.118

Note. "M" denotes the column mean.

Empirical Study

Since the simulation study focused on the loadings and factor inter-correlations, the empirical example presented in the following focused on the robustness of factor score predictors. A sample of 497 German participants (353 females; 71 %; age: $M = 33.1$; $SD = 12.6$) was recruited by means of newspaper advertising and through advertising in university courses. The participants indicated written informed consent and filled in 20 items (10 extraversion items, 10 neuroticism items) of the German Version of the Eysenck Personality Inventory (EPI; Eggert, 1983). Since there are more females in the sample, the data do not represent a balanced sample

of the population. Nevertheless, asymmetries of demographic parameters are not rare in empirical research, so that it seemed reasonable to demonstrate RFT by means of this sample.

Two factors were extracted by means of ULFA, because two factors (Extraversion and Neuroticism) were expected to occur. The Promax-rotated ULFA-solution (Kappa=4), the corresponding Promax-rotated RFT-solution, and the Promax-rotated PCA-solution are presented in Table 6. The Neuroticism-factor is rather clear whereas the Extraversion-factor is rather weak, because four items do not load as expected. Overall, the ULFA loading pattern and the corresponding RFT loading pattern were very similar, although some of the largest ULFA loadings were a bit smaller in the RFT-solution. Moreover, inspection of Table 6 reveals the well-known overestimation of loadings that occurs with PCA.

Table 6. Pattern-loadings of Promax-solution of ULFA, UL-based RFT, and PCA for 20 items of the EPI

item	ULFA		UL-RFT		PCA	
	N	E	N	E	N	E
e01	-.02	.48	-.03	.47	-.04	.58
e03	-.46	.18	-.46	.18	-.52	.22
e05	.17	.28	.18	.32	.20	.40
e08	.16	.37	.16	.40	.17	.50
e10	.38	.18	.39	.19	.44	.24
e13	.00	.51	-.01	.50	-.01	.62
e15	-.16	.44	-.17	.43	-.20	.53
e17	-.05	.57	-.06	.52	-.08	.65
e20	.00	.03	.00	.03	.00	.04
e22	.15	.19	.16	.22	.18	.27
n02	.47	.11	.47	.11	.53	.14
n04	.36	.07	.38	.07	.43	.09
n07	.63	.15	.59	.14	.67	.18
n09	.49	-.07	.49	-.07	.55	-.09
n11	.33	-.16	.35	-.18	.40	-.22
n14	.53	-.03	.53	-.03	.60	-.04
n16	.49	-.08	.49	-.08	.56	-.09
n19	.24	.13	.25	.14	.29	.18
n21	.38	.07	.40	.07	.45	.09
n23	.57	-.06	.55	-.06	.63	-.07
Inter-correlations						
	-.05		-.05		-.03	
First 10 eigenvalues of unrotated PCA:						
3.34, 2.20, 1.40, 1.25, 1.09, 1.02, .96, .89, .86, .79						

RETAINED-COMPONENTS FACTOR TRANSFORMATION

In some occasions researchers want to get scores for each participant, so that Thurstone's regression score predictor was computed for the ULFA-solution and for the RFT-solution (see [Equation 21](#)). Since gender was distributed rather unequally, the robustness of factor score predictors might be questioned. In order to investigate the robustness of the score predictors, 150 random splits of the total sample into two subsamples (249 vs. 248 participants) were performed. The weights for the computation of score predictors were calculated for ULFA, for UL-based RFT, and for PCA component scores in each sub-sample. Then, the weights were applied to compute the score predictors and component score in the total sample so that the root mean squared (RMS) correlation between the scores based on the two sub-samples was computed as an indicator of the robustness of the score predictors. The RMS correlation was .94 with a standard deviation of .06 for ULFA, .97 with a standard deviation of .04 for the UL-RFT score predictors, and .95 with a standard deviation of .07 for PCA.

Conclusion

A transformation of the retained principal component loadings to be as similar as possible to the factor loading matrix was proposed. This transformation was called 'retained-components factor transformation' (RFT). It was shown that Harman's ideal variables factor score predictor based on RFT has perfect correlations with the principal components. It can therefore be concluded that Harman's factor score predictor based on RFT is an optimal summarizer of the observed variables. Moreover, Thurstone's regression score predictor based on RFT was shown to have a perfect multiple correlation with the principal components, indicating that the RFT based regression score predictor summarizes the same overall individual differences as the principal components, even when the variances are distributed differently on the RFT factors and principal components. Thus, the RFT based regression score predictor is also an optimal summarizer of the observed variables.

In a simulation study based on orthogonal and oblique simple structure the means of the population loadings were very similar for FA and RFT. This demonstrates that the RFT-loadings are not inflated as has been found for PCA-loadings when compared to FA-loadings ([Widaman, 1993](#); [Snook & Gorsuch, 1989](#)). Moreover, the RMS difference between the population factor loadings and the sample loadings was overall equal or smaller for RFT-loadings than for FA-loadings and PCA-loadings. This implies that RFT-loadings can be used as estimates of population factor loadings. Moreover, the mean-difference between the RFT-loadings and the population factor loadings was smallest for the RFT

based on UL-estimates. This indicates that the UL-based RFT-loadings might be slightly more precise than other estimates of the factor loadings. Moreover, the underestimation of inter-factor correlations was less pronounced for RFT-solutions than for the FA-solutions.

The empirical example was based on 20 items of the EPI. The simple structure of the two-factor solutions was not perfect and the sample had an unbalanced gender distribution. Thus, the sample contains imperfect data as they occur in empirical research. The Promax loading pattern of the ULFA-solution and the Promax loading pattern of the UL-based RFT-solution were very similar and would probably lead to the same interpretation of the factors whereas the PCA-loadings were again inflated. Nevertheless, many of the largest loadings in the ULFA-solution were smaller in the RFT-solution. The total sample was divided into two-subsamples and the weights for Thurstone's regression score predictor were computed in the subsamples. These weights were then applied to the total-sample in order to compute score predictors. The RMS of the correlation between the score predictors based on sub-sample weights was a bit smaller for ULFA than for the UL-based RFT. This indicates that score predictors that are based on UL-RFT could be a valuable alternative to conventional scores.

To summarize, RFT could be regarded as interesting in several applied settings because the simple structure models investigated in the present simulation study and in the empirical study are relevant for many areas of research. It was found that RFT allows for a model without inflated loadings, which can be used as estimates of population factor loadings. The underestimation of inter-factor correlations was less pronounced when based on RFT than for FA. Moreover, the RFT model implies score predictors that are optimal summarizers of the observed variables and the regression score predictor based on UL-RFT was more robust than the ULFA-based regression score predictor. In this sense, RFT combines the advantages of PCA (score predictors that are optimal summarizers of observed variables) with the advantages of FA (RFT-loadings are not inflated).

It should be noted that the computation of the RFT-loadings can be based on any initial factor model when the analyses are based on the inter-correlations of the observed variables (maximum likelihood, unweighted least-squares, principal axis factoring, etc.). Although RFT was also calculated for MLFA, it should be noted that this is only possible when the analyses are based on the inter-correlations of the observed variables. When covariances are used instead of correlations, MLFA will lead to a scale-free solution whereas PCA will depend on scaling, so that the RFT might be biased. Accordingly, when RFT is based on covariances ULFA or principal axis factoring would be an appropriate method. Moreover, the stability of

the results of the simulation study indicates that the UL-based RFT-loadings should be preferred over ML-based RFT-loadings in small samples. A small R script that can be used in order to compute the RFT-loadings from an initial loading matrix is available from the authors (<http://beauducel.de/research.html>).

References

- Beauducel, A. & Rabe, S. (2009). Model-related factor score predictors for confirmatory factor analysis. *British Journal of Mathematical and Statistical Psychology*, 62, 489-506.
- Bentler, P. M. & De Leeuw, J. (2011). Factor analysis via components analysis. *Psychometrika*, 76, 461-470.
- Conway, J. M. & Huffcutt, A. I. (2003). A review and evaluation of exploratory factor analysis practices in organizational research. *Organizational Research Methods*, 6, 147-168.
- Eggert, D. (1983). *Eysenck-Persönlichkeits-Inventar (EPI)* [Eysenck-Personality Inventory]. Göttingen: Hogrefe.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4, 272-299.
- Grice, J. W. (2001). Computing and evaluation factor scores. *Psychological Methods*, 6, 430-450.
- Guttman, L. (1955). The determinacy of factor score matrices with applications for five other problems of common factor theory. *British Journal of Statistical Psychology*, 8, 65-82.
- Harman, H. H. (1976). *Modern factor analysis* (3rd ed.). Chicago: The University of Chicago Press.
- Harris, R. J. (2001). *A primer of multivariate statistics* (3rd ed.). Hillsdale, NJ: Erlbaum.
- Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *The Journal of Educational Psychology*, 24, 417-441.
- Hurley, J. R. & Cattell, R. B. (1962). The procrustes program: Producing direct rotation to test a hypothesized factor structure. *Behavioral Science*, 7, 258-262.
- Krijnen, W. P. (2006). Some results on mean square error for factor score prediction. *Psychometrika*, 71, 395-409.

- Lawley, D. N. (1940). The estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinburgh*, 60, 64-82.
- Lovie, P. & Lovie, A. D. (1995). The cold equations: Spearman and Wilson on factor indeterminacy. *British Journal of Mathematical and Statistical Psychology*, 48, 237-253.
- Ogasawara, H. (2003). Oblique factors and components with independent clusters. *Psychometrika*, 68, 299-321.
- Preacher, K. J. & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics*, 2, 13-43.
- Sato, M. (1990). Some remarks on principal component analysis as a substitute for factor analysis in monofactor cases. *Journal of the Japanese Statistical Society*, 20, 23-31.
- Schneeweiss, H. (1997). Factors and principal components in the near spherical case. *Multivariate Behavioral Research*, 32, 375-401.
- Schneeweiss, H. & Mathes, H. (1995). Factor analysis and principal components. *Journal of Multivariate Analysis*, 55, 105-124.
- Snook, S. C. & Gorsuch, R. L. (1989). Component analysis versus common factor analysis: A monte carlo study. *Psychological Bulletin*, 106, 148-154.
- Tabachnick, B. G. & Fidell, L. S. (2007). *Using multivariate statistics* (5th Ed.). Boston, MA: Pearson Education.
- Ten Berge, J. M. F. & Kiers, H. A. L. (1997). Are all varieties of PCA the same? A reply to Cadima & Jolliffe. *British Journal of Mathematical and Statistical Psychology*, 50, 367-368.
- Ten Berge, J. M. F. & Knol, D. L. (1985). Scale construction on the basis of components analysis: A comparison of three strategies. *Multivariate Behavioral Research*, 20, 45-55.
- Ten Berge, J. M. F., Krijnen, W. P., Wansbeek, T., & Shapiro, A. (1999). Some new results on correlation-preserving factor scores prediction methods. *Linear Algebra and its Applications*, 289, 311-318.
- Thurstone, L.L. (1935). *The Vectors of Mind*. Chicago: University of Chicago Press.
- Unkel, S. & Trendafilov, N. T. (2010). A majorization algorithm for simultaneous parameter estimation in robust exploratory factor analysis. *Computational Statistics and Data Analysis*, 54, 3348-3358.

RETAINED-COMPONENTS FACTOR TRANSFORMATION

Velicer, W. F. & Jackson, D. N. (1990). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure (with comments and reply). *Multivariate Behavioral Research*, 25, 1-114.

Widaman, K. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters? *Multivariate Behavioral Research*, 28, 263-311.

Robust Winsorized Shrinkage Estimators for Linear Regression Model

Nileshkumar H. Jadhav
D.R.K College of Commerce
Kolhapur, India

Dattatraya N. Kashid
Shivaji University
Kolhapur, India

In multiple linear regression, the ordinary least squares estimator is very sensitive to the presence of multicollinearity and outliers in the response variable. To handle these problems in the data, Winsorized shrinkage estimators are proposed and the performance of these estimators is evaluated through mean square error sense.

Keywords: Multicollinearity, outliers, contaminated normal error, Winsorization, mean square error, multiple linear regression

Introduction

In the multiple linear regression model

$$Y = X\beta + \varepsilon, \quad (1)$$

Y is an vector of n observations on the response variable, X is an $n \times p$ matrix of independent variables known as regressor variables, β is a $p \times 1$ vector of unknown regression parameters and ε is an $n \times 1$ vector of unobserved random errors. Classically, it is assumed that the $\varepsilon_i, i = 1, 2, \dots, n$, are independent and identically normally distributed with zero mean and constant variance σ^2 .

It is well known that when the normality assumption holds, the ordinary least squares (OLS) estimator becomes a maximum likelihood estimator and the best linear unbiased estimator of the unknown regression parameters and has the smallest variance in the class of all linear unbiased estimators. However, the real life data often may not satisfy these assumptions and the violation of assumptions dramatically affects the OLS estimation and consequently the prediction based on the OLS estimator. In the literature, the effect of violation of assumptions has been

Nileshkumar H. Jadhav is an Assistant Professor of Statistics at the D.R.K. College of Commerce. Email at n.nil08@gmail.com. Dr. Kashid is a Professor in the Department of Statistics. Email him at dnk_stats@unishivaji.ac.in.

discussed by many authors (see Birkes and Dodge, 1993; Draper and Smith, 1998; Montgomery, Peck and Vining, 2006).

The near linear dependency between the set of regressor variables produces the problem of multicollinearity in the data. Due to the presence of multicollinearity, the variance of the OLS estimator gets inflated. Consequently, the OLS estimates become unstable and may give misleading results. Various techniques are available in the literature to deal with the problem of multicollinearity. Hoerl and Kennard (1970a, b), Hoerl, Kennard and Baldwin (1975), Liu (1993), Liu (2003) are praiseworthy.

Another important problem that has received considerable attention is the presence of outliers in Y - space. Huber (1973) and Rousseeuw and Leroy (1987) pointed out that the presence of outliers significantly affect the performance of the OLS estimator. In most of the situations, outliers in Y - space are due to heavy tailed distribution of error variable. The least squares fit may be spoiled by small but reasonable deviation from normal error distribution (see Huber, 1973; Andrews, 1974). Many robust parameter estimation methods are available in the literature to handle the problem of outliers in the data.

A simultaneous occurrence of multicollinearity and outliers in Y -space due to non-normality of error variable is considered. To handle the problem of multicollinearity and outliers in the data, a class of Winsorized shrinkage estimators is proposed and the performance is evaluated through estimated mean square error (EMSE). An extensive simulation study was conducted to evaluate the performance of the proposed and existing estimators. Also, a real data example is used to illustrate the performance of the estimators.

Regression Model and Some Estimators

To reduce the notational complexity and lengthy expressions, various authors like Liu (1993), Liu (2003), Montgomery, Peck and Vining (2006), Gao and Liu (2011) used a canonical form of a multiple linear regression model. It is given as

$$Y = Z\alpha + \varepsilon, \quad (2)$$

where $Z = XQ$, $\alpha = Q'\beta$ and $Q = (q_1, q_2, \dots, q_p)$ is an orthogonal matrix of eigenvectors q_1, q_2, \dots, q_p corresponding to eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_p \geq 0$ of $X'X$ matrix. Note that, the use of canonical form does not affect the mean square error (MSE) of the estimator (Liu, 2003).

Some existing estimators were examined to handle the problem of multicollinearity and problem of outliers individually present in the data.

Ordinary Least Squares (OLS) Estimator

It is well known that, when $\varepsilon \sim N(0, \sigma^2 I)$, then the optimal estimator of regression parameters is the OLS estimator. It is denoted by

$$\hat{\alpha}_{OLS} = \Lambda^{-1} Z'Y \quad (3)$$

where $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p)$. It is widely used in regression analysis due to its computational ease. Because the OLS estimator is unbiased, the MSE of $\hat{\alpha}_{OLS}$ is given by

$$\begin{aligned} MSE(\hat{\alpha}_{OLS}) &= \text{tr}(Cov(\hat{\alpha}_{OLS})) \\ &= \sigma^2 \sum_{j=1}^p 1/\lambda_j \end{aligned} \quad (4)$$

where the error variance σ^2 is unknown and estimated by $\hat{\sigma}_{OLS}^2 = (Y - Z\hat{\alpha}_{OLS})'(Y - Z\hat{\alpha}_{OLS}) / (n - p)$.

Ordinary Ridge Regression (ORR) Estimator

To overcome the problem of multicollinearity, several methods are put forwarded in the literature, but the ordinary ridge regression estimator (ORR) proposed by Hoerl and Kennard (1970a, b) is one of the most popular biased estimators for regression parameters. It is defined as

$$\hat{\alpha}_{ORR} = (\Lambda + kI)^{-1} \Lambda \hat{\alpha}_{OLS} \quad (5)$$

where $k > 0$ is a ridge parameter and I is an identity matrix of an order $p \times p$. Because, the ORR estimator is biased, the MSE of ORR estimator is obtained as

$$\begin{aligned} MSE(\hat{\alpha}_{ORR}) &= \text{tr}(Cov(\hat{\alpha}_{ORR})) + (Bias(\hat{\alpha}_{ORR}))' (Bias(\hat{\alpha}_{ORR})) \\ &= \sigma^2 \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k)^2} + k^2 \sum_{j=1}^p \frac{\alpha_j^2}{(\lambda_j + k)^2} \end{aligned} \quad (6)$$

The ridge parameter k plays an important role in minimizing the MSE of the ORR estimator. Various choices for estimator of k are available in the literature, but the estimator proposed by Hoerl, Kennard and Baldwin (1975) is widely used. It is defined as

$$k = \frac{p\hat{\sigma}_{OLS}^2}{\hat{\alpha}'_{OLS}\hat{\alpha}_{OLS}} \quad (7)$$

where $\hat{\sigma}_{OLS}^2$ is the estimate of error variance based on the OLS estimator $\hat{\alpha}_{OLS}$. However, $\hat{\alpha}_{ORR}$ is nonlinear function of k . So, using some of the proposed methods to obtain the value of k becomes complicated.

Liu (LIU) Estimator

Liu (1993) proposed a new biased estimator of α called as LIU estimator and is given by

$$\hat{\alpha}_{LIU} = (\Lambda + I)^{-1} (\Lambda + dI) \hat{\alpha}_{OLS} \quad (8)$$

where $0 < d < 1$, is a Liu parameter. The advantage of the LIU estimator is that $\hat{\alpha}_{LIU}$ is a linear function of d . Therefore, it is easier to choose d in $\hat{\alpha}_{LIU}$ than to choose k in $\hat{\alpha}_{ORR}$. Liu (1993) obtained the MSE of the LIU estimator as

$$\begin{aligned} MSE(\hat{\alpha}_{LIU}) &= tr(Cov(\hat{\alpha}_{LIU})) + (Bias(\hat{\alpha}_{LIU}))' (Bias(\hat{\alpha}_{LIU})) \\ &= \sigma^2 \sum_{j=1}^p \frac{(\lambda_j + d)^2}{\lambda_j (\lambda_j + 1)^2} + (1-d)^2 \sum_{j=1}^p \frac{\alpha_j^2}{(\lambda_j + 1)^2} \end{aligned} \quad (9)$$

where the optimal value of d is

$$d = 1 - \sigma^2 \left[\frac{\sum_{j=1}^p 1 / \lambda_j (\lambda_j + 1)}{\sum_{j=1}^p \alpha_j^2 / (\lambda_j + 1)^2} \right] \quad (10)$$

The unknown parameters α and σ^2 are replaced by their unbiased OLS estimates $\hat{\alpha}_{OLS}$ and $\hat{\sigma}_{OLS}^2$ respectively.

Linearized Ridge Regression (LRR) Estimator

Very recently, Liu and Gao (2011) proposed a linearized ridge regression (LRR) estimator to combat the problem of multicollinearity. It can be expressed as

$$\hat{\alpha}_{LRR} = (\Lambda + I)^{-1} (\Lambda + D) \hat{\alpha}_{OLS} \quad (11)$$

where $D = \text{diag}(d_1, d_2, \dots, d_p)$, $d_j \in \mathbb{R}$, $j = 1, 2, \dots, p$. The optimal value of d_j proposed by Gao and Liu (2011) is given by

$$d_j = \frac{\lambda_j (\alpha_j^2 - \sigma^2)}{\sigma^2 + \lambda_j \alpha_j^2}, j = 1, 2, \dots, p \quad (12)$$

and the unknown quantities α and σ^2 are replaced by their OLS estimates to obtain the estimate of d_j , $j = 1, 2, \dots, p$. Gao and Liu (2011) showed that the LRR estimator attends the lower bound of the MSE of the generalized shrinkage estimators (GSE). The MSE of the LRR estimator is given by (Gao and Liu, 2011)

$$\begin{aligned} \text{MSE}(\hat{\alpha}_{LRR}) &= \text{tr}(\text{Cov}(\hat{\alpha}_{LRR})) + (\text{Bias}(\hat{\alpha}_{LRR}))' (\text{Bias}(\hat{\alpha}_{LRR})) \\ &= \sigma^2 \sum_{j=1}^p \frac{(\lambda_j + d_j)^2}{\lambda_j (\lambda_j + 1)^2} + \sum_{j=1}^p \frac{(1 - d_j)^2 \alpha_j^2}{(\lambda_j + 1)^2} \end{aligned} \quad (13)$$

Here, σ^2 and α are replaced by their suitable estimates $\hat{\sigma}_{OLS}^2$ and $\hat{\alpha}_{OLS}$ respectively to obtain the estimate of the MSE of LRR estimator.

Winsorization Approach

Many robust parameter estimation methods have been developed in the literature to deal with the presence of outliers (see Huber, 1973; Birkes and Dodge, 1993). Winsorization is one of the robust techniques that aim to diminish the effect of outliers in the data. Dixon (1960), Bickel (1965), Dixon and Tukey (1968), Chen and Dixon (1972) discussed this approach. Mutan and Senoglu (2008) noted that the Winsorization does not worsen a good linear relationship on non-contaminated data. Winsorized regression is an effective alternative to the least squares estimation method which reduce the effect of contamination on the regression

coefficient. To illustrate the advantage of Winsorization in estimation of regression coefficients, Yale and Forsythe (1976) introduced various methods of Winsorization and compared with each other and with the OLS estimates. Further study in Winsorization is done by Tan and Tabatabai (1988), Chen Welsh and Chan (2001). A general Winsorization procedure proposed by Yale and Forsythe (1976) is briefly introduced as follows.

Winsorization Methodology

Yale and Forsythe (1976) explained the Winsorization procedure for simple linear regression. It can be easily generalize to the multiple linear regression. In this article, following stepwise algorithm is used to obtain the least squares Winsorized (LSW) estimator. Step 1 to Step 5 are used to obtain least squares Winsorized (LSW) estimator for model given in (1) and Step 6 to Step 8 gives LSW estimator in canonical form of model defined in (2).

Stepwise Algorithm

- Step 1.** Using the model given in (1), obtain the OLS estimates and the predicted values (\hat{Y}_i) of $Y_i, i = 1, 2, \dots, n$.
- Step 2.** Set number of points (g) to be Winsorized at each extreme.
- Step 3.** Obtain the residual values as $r_i = Y_i - \hat{Y}_i$ and order them. Let $r_1 \leq r_2 \leq \dots \leq r_n$ be ordered OLS residuals.
- Step 4.** Obtain the least squares estimator using n observations on Y' and X , where

$$Y'_i = \hat{Y}_i + r'_i$$

and

$$r'_i = \begin{cases} r_{g+1} & i = 1, 2, \dots, g \\ r_i & i = g + 1, \dots, n - g \\ r_{n-g} & i = n - g + 1, \dots, n \end{cases}$$

- Step 5.** Repeat the above Step 4 for fixed number of iteration (b). For each iteration, the baseline data on response variable (Y) has been modified as $Y=Y'$ (after first iteration), $Y=Y''$ (after second iteration), $Y=Y'''$ (after

third iteration) and so on by generating new set of residuals (r' , r'' , r''' and so on).

- Step 6.** The modified dataset at the end of b^{th} iteration is denoted by (Y^*, X) . Standardize the modified dataset in such a way that Y^*X denote the correlation between the modified response variable and the set of regressor variables.
- Step 7.** Convert the standardized modified dataset to canonical form using the matrix of eigenvectors (Q) of $X'X$ matrix.
- Step 8.** Using the canonical form of model, perform the OLS estimation to obtain the LSW estimates of unknown regression parameters.

In this article, 10% and 20% observations are considered for Winsorization ($g = 0.1n, 0.2n$). Nevitt and Tam (1998) conducted a pilot study to decide the number of iterations (b). They found that, after five iterations of data modification, the results shows very little change in parameter estimates. So, five iterations are considered to obtain the LSW estimator. Because, the Winsorization is done only in Y , the diagonal matrix of eigenvalues (Λ) and the corresponding matrix of eigenvectors (Q) of X remains unchanged. Using the canonical form of model given in (2), estimators of unknown regression parameters α are proposed to tackle the problem of multicollinearity and outliers simultaneously in the data.

Proposed Estimators

New estimators based on the LSW estimator are now proposed to handle the simultaneous occurrence of multicollinearity and outliers in the data. The proposed estimators are called as Winsorized shrinkage estimators because they reduce the impact of multicollinearity by shrinking the LSW estimator. The different forms of shrinkage quantity produce the different Winsorized shrinkage estimators. In the following subsections, some Winsorized shrinkage estimators are introduced and their modified MSE Expressions are obtained. The technique suggested by Kan, Alpu and Yazici (2013) is implemented to obtain the modified MSE of the proposed estimators.

Ordinary Ridge Regression Winsorized (ORRW) Estimator

The ordinary ridge regression Winsorized (ORRW) estimator of α , based on the ORR estimator (Hoerl and Kennard, 1970a, b), is defined as

$$\hat{\alpha}_{ORRW} = (\Lambda + k_{LSW} I)^{-1} \Lambda \hat{\alpha}_{LSW} \quad (14)$$

where k_{LSW} is the unknown ridge parameter. It is estimated by using the formula $\hat{k}_{LSW} = p \hat{\sigma}_{LSW}^2 / \hat{\alpha}_{LSW}' \hat{\alpha}_{LSW}$, where p denote the number of regressor variables, the $\hat{\alpha}_{LSW}$ denote the LSW estimator of α and $\hat{\sigma}_{LSW}^2 = (Y - Z \hat{\alpha}_{LSW})' (Y - Z \hat{\alpha}_{LSW}) / (n - p)$ is the estimator of σ^2 based on the LSW estimator. The modified MSE of the ORRW estimator is given by

$$MSE(\hat{\alpha}_{ORRW}) = \sigma^2 \sum_{j=1}^p \frac{\lambda_j}{(\lambda_j + k_{LSW})^2} + k_{LSW}^2 \sum_{j=1}^p \frac{\alpha_j^2}{(\lambda_j + k_{LSW})^2} \quad (15)$$

The unknown parameters σ^2 , α and k_{LSW} are replaced by $\hat{\sigma}_{LSW}^2$, $\hat{\alpha}_{LSW}$ and \hat{k}_{LSW} respectively.

Liu Winsorized (LIUW) Estimator

The Liu Winsorized estimator (LIUW), based on the Liu estimator (Liu, 1993), is defined as

$$\hat{\alpha}_{LIUW} = (\Lambda + I)^{-1} (\Lambda + d_{LSW} I) \hat{\alpha}_{LSW} \quad (16)$$

where d_{LSW} is a Liu parameter and it is obtained by using the following formula

$$d_{LSW} = 1 - \sigma^2 \left[\frac{\sum_{j=1}^p 1 / \lambda_j (\lambda_j + 1)}{\sum_{j=1}^p \alpha_j^2 / (\lambda_j + 1)^2} \right] \quad (17)$$

The estimate of d_{LSW} denoted by \hat{d}_{LSW} is obtained by replacing the unknown parameters σ^2 and α in (17) by their estimates based on the LSW estimator. The modified MSE of LIUW estimator is obtained by

$$MSE(\hat{\alpha}_{LIUW}) = \sigma^2 \sum_{j=1}^p \frac{(\lambda_j + d_{LSW})^2}{\lambda_j (\lambda_j + 1)^2} + (1 - d_{LSW})^2 \sum_{j=1}^p \frac{\alpha_j^2}{(\lambda_j + 1)^2} \quad (18)$$

and the unknown parameters are replaced by their corresponding estimates based on the LSW estimator.

Linearized Ridge Regression Winsorized (LRRW) Estimator

The LRRW estimator based on the LSW estimator, (Liu and Gao, 2011) is defined as

$$\hat{\alpha}_{LRRW} = (\Lambda + I)^{-1} (\Lambda + D_{LSW}) \hat{\alpha}_{LSW} \quad (19)$$

where $(\Lambda + I)^{-1} (\Lambda + D_{LSW})$ is a shrinkage matrix and a diagonal matrix D_{LSW} is an order of $p \times p$ with diagonal elements $d_{LSW_j}, j=1, 2, \dots, p$ such that $d_{LSW_j} \in \mathbb{R}$ is obtained by using the formula

$$d_{LSW_j} = \frac{\lambda_j (\alpha_j^2 - \sigma^2)}{\sigma^2 + \lambda_j \alpha_j^2}, \quad j=1, 2, \dots, p \quad (20)$$

where α and σ^2 are estimated using the LSW estimators $\hat{\alpha}_{LSW}$ and $\hat{\sigma}_{LSW}^2$. The modified MSE of the LRRW estimator is given by

$$MSE(\hat{\alpha}_{LRRW}) = \sigma^2 \sum_{j=1}^p \frac{(\lambda_j + d_{LSW_j})^2}{\lambda_j (\lambda_j + 1)^2} + \sum_{j=1}^p \frac{(1 - d_{LSW_j})^2 \alpha_j^2}{(\lambda_j + 1)^2} \quad (21)$$

Here, σ^2 and α are replaced by their suitable estimates based on the LSW estimator.

Simulation Study

A simulation study was carried out to evaluate the performance of proposed estimators. First, the estimated MSE's (EMSE) of the different estimators are obtained and based on the average EMSE (AEMSE), the existing and proposed

estimators are compared. Secondly, the relative average EMSE's (RAEMSE) of estimators with respect to the OLS estimator are obtained and the average reduction in the estimated MSE's of the estimators with respect to the OLS estimator for the different Winsorization proportions is noted.

Comparison of Estimators through Estimated MSE

The regressor variables are generated using a simulation design proposed by McDonald and Galarneau (1975) as

$$x_{ij} = (1 - \rho^2)^{\frac{1}{2}} \zeta_{ij} + \rho \zeta_{i(p+1)}, \quad i = 1, 2, \dots, n, j = 1, 2, \dots, p \quad (22)$$

where ζ_{ij} are independent pseudo random numbers generated from standard normal distribution and ρ^2 is the correlation between any two regressor variables. The following regression model is used to generate n observations on the response variables

$$Y = 10 + 4X_1 + 6X_2 + 2X_3 + 8X_4 + \varepsilon$$

where the error variable ε is generated using the contaminated normal distribution. The $\delta\%$ contamination is done using the following mixture of normal distributions

$$\varepsilon_i \sim f_{\varepsilon_i}(\cdot) = (1 - \delta) \times N(0, 1) + \delta \times N(0, 10^2).$$

For $\delta = 0\%$, 10% , 20% and 30% , and $n = 20, 30$, and 50 , the different degrees of multicollinearity have been achieved by generating regressor variables using the model given in (22) for $\rho = 0.9, 0.99, 0.999$ and 0.9999 . The $0.1n$ and $0.2n$ points are Winsorized at each extreme to reduce the effect of outlier observations. Hence, the 10% and 20% Winsorized estimators of OLS, ORR, LIU and LRR are denoted by LSW10, ORRW10, LIUW10, LRRW10 and LSW20, ORRW20, LIUW20, LRRW20 respectively.

The EMSE of OLS, ORR, LIU, LRR, OLSW10, ORRW10, LIUW10, LRRW10, OLSW20, ORRW20, LIUW20 and LRRW20 estimators are obtained by replacing the values of unknown parameters with their suitable estimates in their respective MSE expressions. Note that, the EMSE of the LIU, LIUW10 and LIUW20 is considered corresponding to those iterations where the estimate of Liu parameter

(d) lies between 0 and 1. For each combination of sample size (n), degree of multicollinearity (ρ) and contamination proportion (δ), the above simulation experiment is repeated 10,000 times and the AEMSE of these estimators are obtained and reported in Table 1. Also, for sample size $n = 30$, the AEMSE of each estimator was plotted for all combinations of ρ and δ . They are depicted graphically in Figure 1.

Table 1. AEMSE of Estimators

$n = 20$	$\delta = 0\%$				$\delta = 10\%$			
	0.9	0.99	0.999	0.9999	0.9	0.99	0.999	0.9999
OLS	0.0031	0.0254	0.2512	2.4896	0.0315	0.2629	2.6386	26.5099
ORR	0.0030	0.0203	0.1254	1.1043	0.0226	0.1301	1.1880	11.8995
LIU	0.0029	0.0199	0.1804	1.7630	0.0209	0.1525	1.7741	18.4332
LRR	0.0024	0.0133	0.0934	0.8641	0.0153	0.0979	0.9179	9.2182
LSW10	0.0020	0.0169	0.1671	1.6480	0.0081	0.0676	0.6723	6.7148
ORRW10	0.0020	0.0144	0.0979	0.8729	0.0072	0.0476	0.4125	4.0837
LIUW10	0.0020	0.0138	0.1219	1.1870	0.0069	0.0500	0.5392	5.5197
LRRW10	0.0017	0.0098	0.0727	0.6783	0.0053	0.0349	0.3179	3.1419
LSW20	0.0013	0.0105	0.1045	1.0304	0.0043	0.0354	0.3482	3.4815
ORRW20	0.0013	0.0095	0.0704	0.6329	0.0040	0.0291	0.2608	2.5935
LIUW20	0.0013	0.0091	0.0789	0.7568	0.0039	0.0292	0.2931	2.9764
LRRW20	0.0011	0.0068	0.0527	0.4941	0.0032	0.0221	0.2048	2.0340
	$\delta = 20\%$				$\delta = 30\%$			
	0.9	0.99	0.999	0.9999	0.9	0.99	0.999	0.9999
OLS	0.0584	0.4960	4.8625	48.0399	0.0849	0.7186	7.0523	69.7408
ORR	0.0383	0.2290	2.1818	21.7413	0.0526	0.3279	3.1753	31.3082
LIU	0.0371	0.3185	3.3957	34.1454	0.0550	0.4875	4.9287	48.9810
LRR	0.0261	0.1766	1.6910	16.7676	0.0363	0.2545	2.4522	24.1812
LSW10	0.0193	0.1598	1.5703	15.5570	0.0345	0.2865	2.8498	28.0075
ORRW10	0.0156	0.0998	0.9268	9.1832	0.0263	0.1691	1.6315	16.0833
LIUW10	0.0148	0.1131	1.2099	12.2463	0.0255	0.2045	2.1224	20.8768
LRRW10	0.0112	0.0754	0.7156	7.0553	0.0185	0.1298	1.2524	12.3462
LSW20	0.0092	0.0755	0.7412	7.2817	0.0162	0.1347	1.3280	12.9899
ORRW20	0.0082	0.0582	0.5489	5.4090	0.0140	0.0989	0.9536	9.4040
LIUW20	0.0080	0.0606	0.6169	6.0481	0.0136	0.1064	1.0715	10.4862
LRRW20	0.0063	0.0448	0.4308	4.2137	0.0104	0.0768	0.7429	7.3156

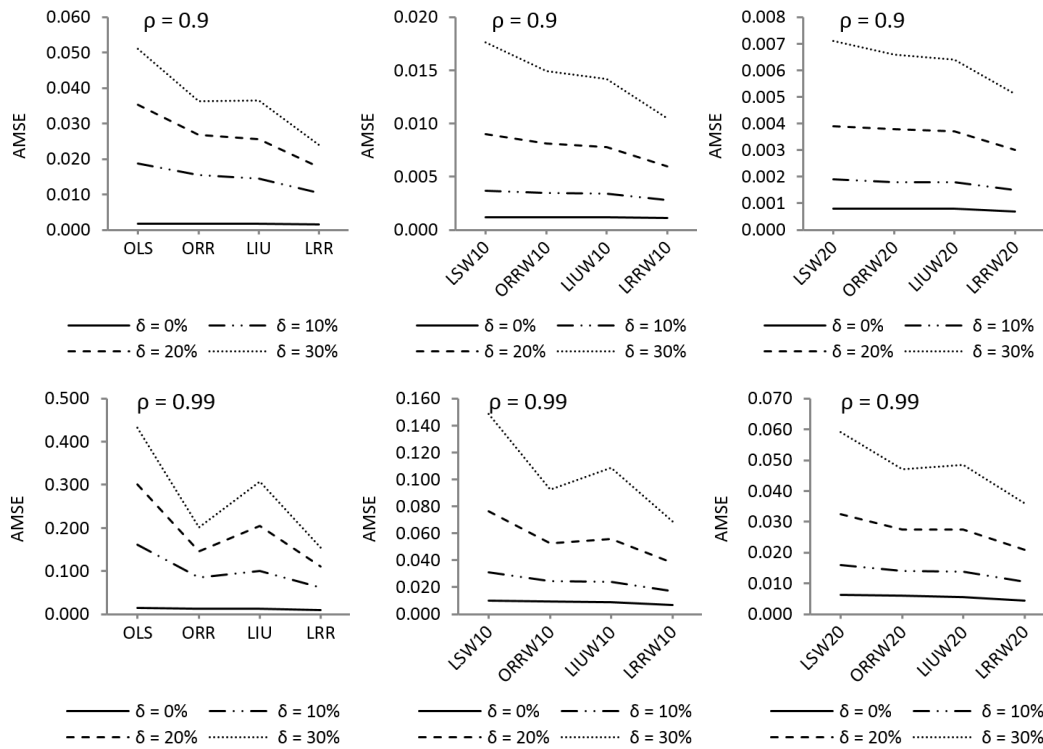
ROBUST WINSORIZED SHRINKAGE ESTIMATORS

Table 1, continued.

<i>n</i> = 30	$\delta = 0\%$				$\delta = 10\%$			
	0.9	0.99	0.999	0.9999	0.9	0.99	0.999	0.9999
OLS	0.0018	0.0152	0.1504	1.5042	0.0188	0.1618	1.5641	15.4151
ORR	0.0018	0.0133	0.0829	0.6727	0.0154	0.0851	0.6966	6.9166
LIU	0.0018	0.0126	0.1122	1.1217	0.0145	0.1003	1.0987	11.5955
LRR	0.0015	0.0091	0.0586	0.5276	0.0104	0.0619	0.5456	5.3711
LSW10	0.0012	0.0102	0.1008	1.0052	0.0037	0.0309	0.2943	2.9777
ORRW10	0.0012	0.0093	0.0646	0.5373	0.0035	0.0244	0.1856	1.8383
LIUW10	0.0012	0.0088	0.0763	0.7571	0.0034	0.0241	0.2343	2.4833
LRRW10	0.0011	0.0066	0.0457	0.4190	0.0028	0.0173	0.1428	1.4246
LSW20	0.0008	0.0063	0.0621	0.6184	0.0019	0.0159	0.1519	1.5370
ORRW20	0.0008	0.0059	0.0455	0.3881	0.0018	0.0141	0.1177	1.1710
LIUW20	0.0008	0.0056	0.0486	0.4768	0.0018	0.0137	0.1290	1.3243
LRRW20	0.0007	0.0044	0.0327	0.3036	0.0015	0.0106	0.0924	0.9255
	$\delta = 20\%$				$\delta = 30\%$			
	0.9	0.99	0.999	0.9999	0.9	0.99	0.999	0.9999
OLS	0.0354	0.3013	2.9141	29.3682	0.0511	0.4325	4.1930	42.7257
ORR	0.0268	0.1462	1.3014	12.9487	0.0363	0.2013	1.8649	19.0030
LIU	0.0256	0.2055	2.1456	21.7532	0.0365	0.3080	3.1180	31.5423
LRR	0.0177	0.1105	1.0168	10.0854	0.0239	0.1550	1.4581	14.8004
LSW10	0.0090	0.0761	0.7419	7.4299	0.0176	0.1485	1.4250	14.6328
ORRW10	0.0081	0.0523	0.4451	4.4012	0.0149	0.0924	0.8218	8.4375
LIUW10	0.0078	0.0559	0.5759	5.8704	0.0142	0.1084	1.1069	11.2143
LRRW10	0.0060	0.0383	0.3453	3.4063	0.0105	0.0690	0.6388	6.5194
LSW20	0.0039	0.0325	0.3183	3.1710	0.0071	0.0592	0.5764	5.8364
ORRW20	0.0038	0.0274	0.2455	2.4276	0.0066	0.0470	0.4311	4.3970
LIUW20	0.0037	0.0274	0.2684	2.6864	0.0064	0.0484	0.4791	4.8429
LRRW20	0.0030	0.0209	0.1938	1.9226	0.0051	0.0359	0.3402	3.4536
<i>n</i> = 50	$\delta = 0\%$				$\delta = 10\%$			
	0.9	0.99	0.999	0.9999	0.9	0.99	0.999	0.9999
OLS	0.0010	0.0084	0.0832	0.8325	0.0107	0.0893	0.8805	8.6261
ORR	0.0010	0.0078	0.0536	0.3722	0.0097	0.0540	0.3981	3.8131
LIU	0.0010	0.0074	0.0646	0.6398	0.0092	0.0606	0.6416	6.5740
LRR	0.0009	0.0057	0.0351	0.2932	0.0069	0.0366	0.3126	2.9891
LSW10	0.0007	0.0057	0.0560	0.5618	0.0015	0.0124	0.1236	1.2125
ORRW10	0.0007	0.0054	0.0408	0.3007	0.0015	0.0111	0.0847	0.7510
LIUW10	0.0007	0.0052	0.0442	0.4342	0.0015	0.0105	0.0996	0.9978
LRRW10	0.0006	0.0041	0.0273	0.2347	0.0013	0.0079	0.0628	0.5858
LSW20	0.0004	0.0035	0.0344	0.3439	0.0008	0.0066	0.0655	0.6442
ORRW20	0.0004	0.0034	0.0279	0.2184	0.0008	0.0062	0.0539	0.4973
LIUW20	0.0004	0.0033	0.0279	0.2713	0.0008	0.0060	0.0560	0.5517
LRRW20	0.0004	0.0027	0.0193	0.1703	0.0007	0.0048	0.0415	0.3962

Table 1, continued.

$n = 50$	$\delta = 20\%$				$\delta = 30\%$			
	0.9	0.99	0.999	0.9999	0.9	0.99	0.999	0.9999
OLS	0.0200	0.1674	1.6402	16.3004	0.0288	0.2411	2.3471	23.6339
ORR	0.0170	0.0894	0.7280	7.1945	0.0234	0.1218	1.0353	10.4008
LIU	0.0161	0.1205	1.2570	12.4734	0.0224	0.1807	1.8062	18.3140
LRR	0.0114	0.0638	0.5749	5.6323	0.0153	0.0907	0.8127	8.1273
LSW10	0.0038	0.0314	0.3116	3.0516	0.0082	0.0682	0.6680	6.7232
ORRW10	0.0037	0.0249	0.1946	1.8493	0.0076	0.0480	0.3886	3.8843
LIUW10	0.0036	0.0245	0.2442	2.4453	0.0073	0.0513	0.5133	5.2782
LRRW10	0.0029	0.0173	0.1497	1.4394	0.0056	0.0342	0.3015	3.0003
LSW20	0.0015	0.0126	0.1242	1.2294	0.0028	0.0230	0.2242	2.2560
ORRW20	0.0015	0.0115	0.1006	0.9742	0.0027	0.0201	0.1762	1.7555
LIUW20	0.0015	0.0111	0.1067	1.0578	0.0027	0.0197	0.1902	1.9220
LRRW20	0.0013	0.0087	0.0795	0.7768	0.0022	0.0152	0.1390	1.3828

Figure 1. AEMSE plot of various estimators for different combinations of ρ and δ

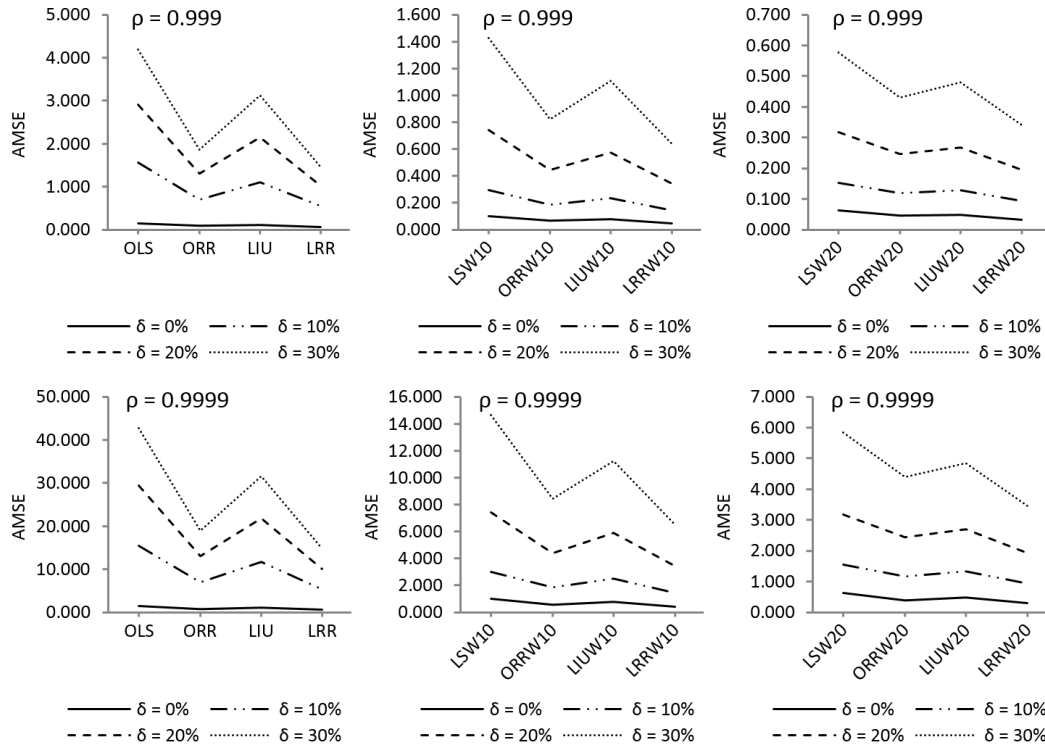


Figure 1, continued.

Table 1 and Figure 1 clearly indicate that

- For each combination of n , ρ and δ , the LRRW20 estimator has consistently smaller AEMSE value than that of the other estimators. It clearly indicates that the estimator LRRW20 shows better performance as compare to the other estimators in the EMSE sense.
- The AEMSE of each estimator decreases with increase in sample size (n), but it increases with increase in the proportion of contamination (δ) in the error variable.
- When degree of multicollinearity increases, the AEMSE of each estimator is also increases.
- For any combination of n , ρ and δ , as degree of Winsorization (δ) increases, the AEMSE of each estimator goes on decreases.

Relative AEMSE (RAEMSE) comparison

The RAEMSE is one of the suitable measure to evaluate the performance of estimators. The RAEMSE of estimator 'T' with respect to the OLS estimator is obtained by using the formula

$$RAEMSE_T = (AEMSE_{OLS} - AEMSE_T) / AEMSE_{OLS}$$

where $AEMSE_{OLS}$ and $AEMSE_T$ denote the AEMSE of the OLS estimator and considered estimator 'T'. The maximum value of $RAEMSE_T$ is one. $RAEMSE_T$ greater than zero indicates the corresponding estimator 'T' performs better than the OLS estimator in AEMSE sense. The $RAEMSE_T$ close to one indicates the corresponding estimator 'T' outperforms as compare to the OLS estimator.

Using the AEMSE's of the OLS, ORR, LIU, LRR, OLSW10, ORRW10, LIUW10, LRRW10, OLSW20, ORRW20, LIUW20 and LRRW20 estimators obtained in Table 1, the RAEMSE of each estimator was computed with respect to the OLS estimator. For all combinations of ρ and δ with $n = 30$, the RAEMSE of each estimator is presented in Table 2. Also, RAEMSE of all considered estimators with respect to the OLS estimator is plotted in Figure 2.

Table 2. RAEMSE of estimators with the OLS estimator for $n = 30$

	$\rho = 0.9$				$\rho = 0.99$			
	$\delta = 0\%$	$\delta = 10\%$	$\delta = 20\%$	$\delta = 30\%$	$\delta = 0\%$	$\delta = 10\%$	$\delta = 20\%$	$\delta = 30\%$
ORR	0.0000	0.1809	0.2429	0.2896	0.1250	0.4740	0.5148	0.5346
LIU	0.0000	0.2287	0.2768	0.2857	0.1711	0.3801	0.3180	0.2879
LRR	0.1667	0.4468	0.5000	0.5323	0.4013	0.6174	0.6333	0.6416
LSW10	0.3333	0.8032	0.7458	0.6556	0.3289	0.8090	0.7474	0.6566
ORRW10	0.3333	0.8138	0.7712	0.7084	0.3882	0.8492	0.8264	0.7864
LIUW10	0.3333	0.8191	0.7797	0.7221	0.4211	0.8511	0.8145	0.7494
LRRW10	0.3889	0.8511	0.8305	0.7945	0.5658	0.8931	0.8729	0.8405
LSW20	0.5556	0.8989	0.8898	0.8611	0.5855	0.9017	0.8921	0.8631
ORRW20	0.5556	0.9043	0.8927	0.8708	0.6118	0.9129	0.9091	0.8913
LIUW20	0.5556	0.9043	0.8955	0.8748	0.6316	0.9153	0.9091	0.8881
LRRW20	0.6111	0.9202	0.9153	0.9002	0.7105	0.9345	0.9306	0.9170

ROBUST WINSORIZED SHRINKAGE ESTIMATORS

Table 2, continued.

	$\rho = 0.999$				$\rho = 0.9999$			
	$\delta = 0\%$	$\delta = 10\%$	$\delta = 20\%$	$\delta = 30\%$	$\delta = 0\%$	$\delta = 10\%$	$\delta = 20\%$	$\delta = 30\%$
ORR	0.4488	0.5546	0.5534	0.5552	0.5528	0.5513	0.5591	0.5552
LIU	0.2540	0.2976	0.2637	0.2564	0.2543	0.2478	0.2593	0.2617
LRR	0.6104	0.6512	0.6511	0.6523	0.6492	0.6516	0.6566	0.6536
LSW10	0.3298	0.8118	0.7454	0.6601	0.3317	0.8068	0.7470	0.6575
ORRW10	0.5705	0.8813	0.8473	0.8040	0.6428	0.8807	0.8501	0.8025
LIUW10	0.4927	0.8502	0.8024	0.7360	0.4967	0.8389	0.8001	0.7375
LRRW10	0.6961	0.9087	0.8815	0.8477	0.7214	0.9076	0.8840	0.8474
LSW20	0.5871	0.9029	0.8908	0.8625	0.5889	0.9003	0.8920	0.8634
ORRW20	0.6975	0.9247	0.9158	0.8972	0.7420	0.9240	0.9173	0.8971
LIUW20	0.6769	0.9175	0.9079	0.8857	0.6830	0.9141	0.9085	0.8867
LRRW20	0.7826	0.9409	0.9335	0.9189	0.7982	0.9400	0.9345	0.9192

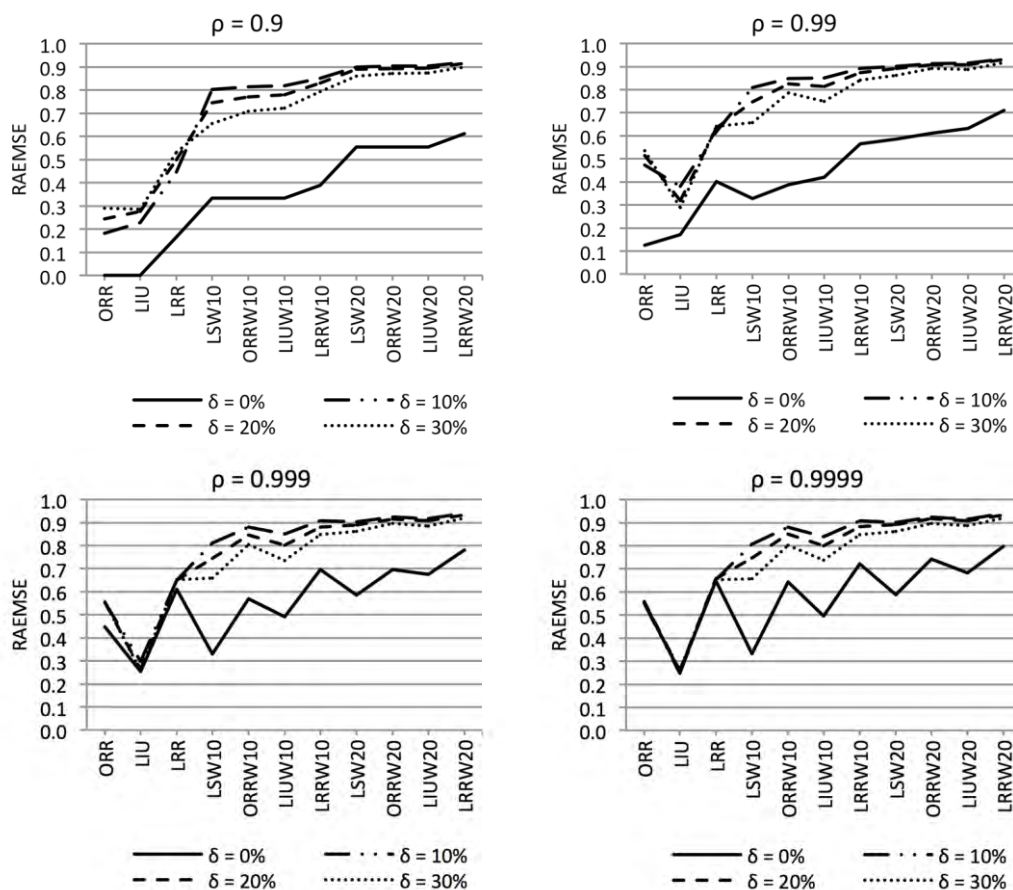


Figure 2. Line plot plot of RAEMSE of estimators with respect to the OLS estimator

Table 2 and Figure 2 show that

- For 0% contamination, as degree of multicollinearity (ρ) increases, the RAEMSE of each estimator with respect to the OLS estimator is also increases.
- With 0% contamination and for $\rho = 0.9$, on an average, 10% Winsorized shrinkage estimators (LSW10, ORRW10, LIUW10 and LRRW10) shows 34.72% reduction in AEMSE with respect to the OLS estimator. Similarly, for $\rho = 0.99, 0.999$ and 0.9999 , it shows 42.60%, 52.23% and 54.82% reduction respectively. Also for $\rho = 0.9, 0.99, 0.999$ and 0.9999 , the 20% Winsorized shrinkage estimators (LSW20, ORRW20, LIUW20 and LRRW20), on an average shows 56.94%, 63.49%, 68.60% and 70.30% reduction in AEMSE respectively.
- On the similar line, for $\delta = 30\%$, the 10% Winsorization shows on an average 72.02%, 75.82%, 76.20% and 76.12% reduction in AEMSE for $\rho = 0.9, 0.99, 0.999$ and 0.9999 and for 20% Winsorization, it is 87.67%, 88.99%, 89.11% and 89.16% respectively.

Real Data Example

A real data set on tobacco blends given by Myers (1990) is used to evaluate the performance of various estimators. The response variable Y measures the heat evolved from tobacco during the smoking process. This data set contains 30 observations and four regressor variables namely X_1, X_2, X_3 , and X_4 . Arslan and Billor (2000) noted that the tobacco blends data suffers from the problem of multicollinearity and outliers simultaneously. The variance inflation factor (VIF) values for each term are 324.1412, 45.1728, 173.2577 and 138.1753. It indicates the severe problem of multicollinearity.

For this real data, the estimate of the bias (EBIAS), variance (EVAR) and MSE (EMSE) of the OLS, ORR, LIU, LRR, OLSW10, ORRW10, LIUW10, LRRW10, OLSW20, ORRW20, LIUW20 and LRRW20 estimators were obtained and are reported in Table 3. Also, the relative EMSE (REMSE) of each estimator with respect to the OLS estimator is computed and presented in Table 3. Positive value of REMSE implies the performance of the corresponding estimator is better than the OLS estimator.

ROBUST WINSORIZED SHRINKAGE ESTIMATORS

Table 3. EBIAS, EVAR, EMSE and REMSE of Estimators

Estimators	EBIAS	EVAR	EMSE	REMSE	REMSE (in %)
OLS	0.000000	1.120600	1.120600	-	-
ORR	0.352000	0.482000	0.937300	0.163573	16.357300
LIU	-0.058100	0.544200	0.883900	0.211226	21.122600
LRR	0.078300	0.647300	0.850400	0.241121	24.112100
LSW10	0.000000	0.607500	0.607500	0.457880	45.788000
ORRW10	0.125700	0.325700	0.480600	0.571123	57.112300
LIUW10	-0.157600	0.382400	0.507900	0.546761	54.676100
LRRW10	0.120200	0.367800	0.469400	0.581117	58.111700
LSW20	0.000000	0.088100	0.088100	0.921381	92.138100
ORRW20	0.045000	0.080400	0.086700	0.922631	92.263100
LIUW20	-0.012700	0.083900	0.086000	0.923255	92.325500
LRRW20	0.018600	0.083200	0.085600	0.923612	92.361200

From Table 3, it seems that the increase in Winsorization proportion reduces the EVAR and EMSE of each estimator. 10% and 20% Winsorization on an average shows 53.92% and 92.27% reduction in the EMSE with respect to the OLS estimator respectively. Also, LRRW20 shows smaller EMSE as compare to other estimators.

Conclusion

A Winsorized form of the OLS estimator, ORR estimator, LIU estimator and LRR estimators are introduced. A simulation study and a real data example show that the Winsorization procedure reduces the EMSE of estimators and improve the performance of the estimators. Also, it reveals that the LRR estimator with 20% Winsorization shows consistently minimum EMSE among the all other considered estimators.

References

- Andrews, D. F. (1974). A Robust Method for Multiple Linear Regression. *Technometrics*, 16(4), 523-531.
- Arslan, O., & Billor, N. (2000). Robust Liu estimator for regression based on an M-estimator. *Journal of Applied Statistics*, 27(1), 39-47.
- Bickel, P. J. (1965). On Some Robust Estimates of Location. *Annals of Mathematical Statistics*, 36, 847-858.
- Birkes, D., & Dodge, Y. (1993). *Alternative methods of regression*. New York: John Wiley & Sons.
- Chen, E. H., & Dixon, W. J. (1972). Estimates of Parameters of a Censored Regression Sample. *Journal of the American Statistical Association*, 67, 664-671.
- Chen, L. A., Welsh, A. H., & Chan, W. (2001). Estimators for the Linear Regression Model based on Winsorized Observations. *Statistica Sinica*, 11, 147-172.
- Dixon, W. J. (1960). Simplified Estimation from Censored Normal Samples. *Annals of Mathematical Statistics*, 3, 385-391.
- Dixon, W. J., & Tukey, J. W. (1968). Approximate Behaviour of the Distribution of Winsorized t (Trimming / Winsorization 2). *Technometrics*, 10, 83-98.
- Draper, N. R., & Smith, H. (1998). *Applied Regression Analysis* (3rd Ed.). New York: John Wiley & Sons.
- Gao, F., & Liu, X. Q. (2011). Linearized Ridge Regression Estimator under the Mean Square Error Criterion in a Linear Regression Model. *Communications in Statistics – Simulation and Computation*, 40, 1434-1443.
- Hoerl, A. E., & Kennard, R. W. (1970a). Ridge regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12, 55-67.
- Hoerl, A. E., & Kennard, R. W. (1970b). Ridge regression: Applications to Nonorthogonal Problems. *Technometrics*, 12, 69-82.
- Hoerl, A. E., Kennard, R. W., & Baldwin, K. F. (1975). Ridge regression: Some Simulations. *Communications in Statistics – Theory and Methods*, 4(2), 105-123.
- Huber, P. J. (1973). Robust Regression: Asymptotics, Conjectures, and Monte Carlo. *Annals of Statistics*, 1, 799-821.

ROBUST WINSORIZED SHRINKAGE ESTIMATORS

- Kan, B., Alpu, O., & Yazici, B. (2013). Robust ridge and robust Liu estimator for regression based on the LTS estimator. *Journal of Applied Statistics*, 40(3), 644-655.
- Liu, K. (1993). A New Class of Biased Estimate in Linear Regression. *Communications in Statistics – Theory and Methods*, 22, 393-402.
- Liu, K. (2003). Using Liu-Type Estimator to Combat Collinearity. *Communications in Statistics – Theory and Methods*, 32(5), 1009-1020.
- Liu, X. Q., & Gao, F., (2011). Linearized Ridge Regression Estimator in Linear Regression. *Communications in Statistics – Theory and Methods*, 40(12), 2182-2192.
- McDonald, G. C., & Galarneau, D. I. (1975). A Monte Carlo Evaluation of Some Ridge-type Estimators. *Journal of the American Statistical Association*, 70(350), 407-416.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2006). *Introduction to Linear Regression Analysis* (3rd ed.). New York: John Wiley & Sons.
- Mutan, O. C., & Senoglu, B. (2008). A Monte Carlo Comparison of Regression Estimators When the Error Distribution is Long-Tailed Symmetric. *Journal of Modern Applied Statistical Methods*, 8(1), 161-172.
<http://digitalcommons.wayne.edu/jmasm/vol8/iss1/14/>
- Myers, R. H. (1990). *Classical and Modern Regression with Applications* (2nd ed.). Boston: Duxbury.
- Nevitt, J., & Tam, H. P. (1998). A Comparison of Robust and Nonparametric Estimators under the Simple Linear Regression Model. *Multiple Regression Viewpoints*, 25, 54-69.
- Rousseeuw, P. J., & Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. New York: John Wiley & Sons.
- Tan, W. Y., & Tabatabai, M. A. (1988). A Modified Winsorized Regression Procedure for Linear Models. *Journal of Statistical Computation and Simulation*, 30, 299-313.
- Yale, C., & Forsythe, A. B. (1976). Winsorized Regression. *Technometrics*, 18, 291-300.

Pairwise Comparison in Repeated Measures

I. C. A. Oyeka

Nnamdi Azikiwe University
Awka, Nigeria

C. C. Nnanatu

Nnamdi Azikiwe University
Awka, Nigeria

Sometimes a random sample of subjects or patients may be exposed to a battery of diagnostic tests or medication over time and interest is on determining whether there is progressive remission of condition, disease or symptom. Also perhaps early in a program or experiment, subjects or candidates may be required to significantly improve in their performance rates at the current trial relative to an immediately preceding trial, otherwise they may have to withdraw from or drop out. The research interest would then be to determine some critical minimum marginal success rate to guide the management in decision making as well as in policy implementation. Success rates lower than the minimum expected value would indicate a need for some remedial actions. A method of estimating these rates is proposed assuming that the requirement is at the second trial. Pairwise comparisons of proportions of success or failure by subjects or candidates in a sequence of experiments or trials over time or space are conducted to ascertain which subject or combinations is responsible for the rejection of the null hypothesis. The proposed methods is illustrated and shown to be at least as efficient and powerful as competitors.

Keywords: Repeated measures, data coding, pairwise comparison

Introduction

Sometimes a researcher may obtain repeated measurements or responses on objects, subjects or items, often measured on an ordinal scale over space or over a number of time periods or set of treatment conditions, experiments or trials. The subjects or candidates are considered random but the treatment conditions may be either fixed or also random. These data may not be available in numerical form, but in the form of letter grades earned by candidates at examinations, continuous assessments or job interviews, grades (or grade point averages) earned at the end of each year during students' studies in an educational institution, etc. Thus each subject or candidate is exposed to each of the experimental conditions over time or space.

I.C.A. Oyeka is a professor of statistics in the Department of Statistics at Nnamdi Azikiwe University; Email him at cyprainoyeka@yahoo.com. C.C. Nnanatu is a statistics lecturer at Nnamdi Azikiwe University. Email him at nnanatuchibuzor@yahoo.com.

PAIRWISE COMPARISON IN REPEATED MEASURES

At each of these instances, the interest may be in determining whether the subjects improved their performances or chances of success over the set of treatment conditions during the study or experimental period. That is, interest is in testing whether the proportions of positive responses are the same or different over a set of treatment conditions.

If the null hypothesis of no improvement is rejected in which case, there perhaps exists some improvements in performance or increases in proportions of positive responses, one may then proceed further to statistically examine any observed patterns in these increases, with a view to ascertaining which of the treatment conditions or their combinations may have led to a rejection of the null hypothesis of equal experimental conditions or treatment proportions of success. Often interest in these regards may be in determining whether the subjects on the average successively improve their performance rather than in multiple comparisons of all the treatment conditions. That is, research interest may be in pairwise comparisons of proportions of success or failure by subjects or candidates in a sequence of experiments or trials over time or space.

Several nonparametric methods exist for answering these questions. For example one may rank order the observations for each subject or candidate across the treatment conditions and then apply any of the non-parametric methods used in analyzing ordered data (see for example Conover, 1980; Friedman, 1937; Oyeka, 1986; Page, 1963; Prentice, 1978; Puri and Sen, 1967; Kempthorne, 1979).

The Proposed Method

A method is proposed based on an earlier work by Oyeka (1990). Assume that there are r independently drawn subjects candidates involved in the study each of whom is observed and scored at each of time periods, location or treatment conditions.

Specifically let x_{ij} be the score or grade earned by the i^{th} subject at the j^{th} treatment condition or trial for $i = 1, 2, \dots, r$ and $j = 1, 2, \dots, c$. The data format is as in Table 1.

Now let

$$u_{ij} = \begin{cases} 1, & \text{if } x_{ij} > x_{i,j-1} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

For $i = 1, 2, \dots, r$ and $j = 2, 3, \dots, c$

Thus the i^{th} subject or candidate is assigned a score of 1 if the subject's score or grade of the current trial, interview, examination, experiment or location is

higher than the subject score of the immediately preceding trial; otherwise the candidate is assigned a score of 0.

Table 1. Data format for the scores or grades earned by the i^{th} subject at the j^{th} treatment condition.

Subject	Treatment				
	1	2	3	...	c
1	X_{11}	X_{12}	X_{13}	...	X_{1c}
2	X_{21}	X_{22}	X_{23}	...	X_{2c}
3	X_{31}	X_{32}	X_{33}	...	X_{3c}
\vdots	\vdots	\vdots	\vdots		\vdots
r	X_{r1}	X_{r2}	X_{r3}	...	X_{rc}

This coding scheme is also appropriate if interest is in comparing one treatment, location or period (control) with the others. In this case an observation in any of the other treatments is coded if it is greater than the corresponding observation in the control treatment, otherwise it is coded 0.

Let

$$\pi_j = p(u_j = 1) \quad (2)$$

and

$$t_j = \sum_{i=1}^r u_{ij} \quad (3)$$

Note that t_j is the number of ones (1's) or successes by subjects in the current trial relative to the immediately preceding trial. The corresponding number of zeroes (0's) or failures is $r - t_j$. Let

$$n_{1.} = t = \sum_{j=2}^c t_j \quad (4)$$

be the total number of 1's for all the c experimental conditions and let

$$n_{2.} = \sum_{j=2}^c r - t_j = r(c-1) - t \quad (5)$$

PAIRWISE COMPARISON IN REPEATED MEASURES

be the total number of 0's or failures for all the c treatment conditions.

Hence,

$$E[u_{ij}] = \pi_j; \text{Var}(u_{ij}) = \pi_j(1 - \pi_j) \quad (6)$$

and

$$E[t_j] = r\pi_j; \text{Var}(t_j) = r\pi_j(1 - \pi_j) \quad (7)$$

Note that π_j is the proportion of 1's or successes in the current j^{th} trial relative to the immediately preceding $(j - 1)^{\text{th}}$ trial and is estimated as

$$p_j = \frac{t_j}{r} \quad (8)$$

and the corresponding variance is estimated as

$$\text{Var}(p_j) = \frac{\hat{\pi}_j(1 - \hat{\pi}_j)}{r} = \frac{t_j(r - t_j)}{r^3} \quad (9)$$

If the proportions of successes are the same for all the c trials or treatments then the common proportion may be estimated as

$$p = \frac{t}{r(c-1)} \quad (10)$$

These results may be summarized in a $2 \times (c - 1)$ table, as in [Table 2](#).

The observed number of 1's or successes and number of 0's or failures for the j^{th} treatment condition are respectively

$$O_{1j} = t_j \text{ and } O_{2j} = r - t_j \quad (11)$$

Table 2. $2 \times (c - 1)$ Table for Analysis for Repeated Measures.

<i>Observations</i>	<i>Treatment Conditions</i>				Total
	2	3	...	c	
Number of 1's (t_j)	t_2	t_3	...	t_c	t
Number of 0's ($r - t_j$)	$r - t_2$	$r - t_3$...	$r - t_c$	$r(c - 1) - t$
Total	r	r	...	r	$r(c - 1)$
Proportion (p_j)	p_2	p_3	...	p_c	$p = \frac{t}{r(c - 1)}$

Under the Null hypothesis of equal population proportions of successes for all the treatment or experimental conditions, the corresponding expected numbers of 1's (successes) and 0's (failures) are respectively

$$E_{1j} = \frac{rt}{r(c-1)} \text{ and } E_{2j} = \frac{r(r(c-1)-t)}{r(c-1)} \quad (12)$$

and under the null hypothesis of no difference between treatments or periods in the success rates achieved, the test statistic

$$\chi^2 = \sum_{i=1}^2 \sum_{j=2}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (13)$$

has approximately a chi-square distribution with $c - 1$ degrees of freedom and may be used to test the null hypothesis of no difference in success rates.

Now using Equations (11) and (12)

$$\chi^2 = \sum_{j=2}^c \left(\frac{(t_j - rt)^2}{\frac{r(c-1)}{rt}} \right) + \sum_{j=2}^c (r - t_j) - \frac{\frac{r(r(c-1)-t)^2}{r(c-1)}}{\frac{r(r(c-1)-t)}{r(c-1)}} \quad (14)$$

which when simplified yields

PAIRWISE COMPARISON IN REPEATED MEASURES

$$\chi^2 = \frac{r(c-1)^2}{t(r(c-1)-t)} \sum_{j=2}^c \frac{(t_j - t)^2}{(c-1)} \quad (15)$$

which has approximately a chi-square distribution with $c - 1$ degrees of freedom.

An equivalent expression for Equation (13) in terms of the proportions in Equations (8) and (10) is

$$\chi^2 = \frac{r \sum_{j=2}^c (p_j - p)^2}{pq} \quad (16)$$

and may be used to test the null hypotheses of equal population proportion success provided $rc \geq 30$, where $q = 1 - p$.

The null hypothesis, H_0 is rejected at the α level of significance if

$$\chi^2 \geq \chi_{1-\alpha; c-1}^2 \quad (17)$$

otherwise H_0 is accepted.

If H_0 is rejected then one may wish to proceed to investigate further which treatments, experimental conditions or their combinations may have led to the rejection of the null hypothesis. In particular, one may wish to test whether subjects are successively improving their performance over time, space or experimental conditions.

Now let π_j and π_k as defined above be the population proportions of positive responses or successes at the j^{th} and k^{th} trials respectively for $j, k = 2, 3, \dots, c, j \neq k$, with the corresponding sample estimates of p_j and p_k respectively.

$$p_j = \frac{t_j}{r} \text{ and } p_k = \frac{t_k}{r} \text{ as in equation (8).}$$

Note that π_j and π_k respectively measure the percentage increases in performance by a population of subjects at the j^{th} and k^{th} treatment conditions relative to their performance at the $(j-1)^{\text{th}}$ and $(k-1)^{\text{th}}$ treatment conditions respectively. Interest may then be in testing either the null hypothesis that these relative improvement rates differ by some constants or the null hypothesis that each of the relative difference equal to some constant or the null hypothesis that there is

no relative improvement. Notationally, these null hypotheses may be expressed as either

$$1. \quad H_0: \pi_j - \pi_k \geq \pi_0 \text{ versus } H_1: \pi_j - \pi_k < \pi_0, \text{ say, } j \neq k, (-1 < \pi_0 < 1) \quad (18)$$

$$2. \quad H_0: \pi_j - \pi_0 \text{ versus } H_1: \pi_j < \pi_0, \text{ say } j = 2, 3, \dots, c \quad (19)$$

To test the null hypothesis of equation (18), note that the sample estimate of $\pi_j - \pi_k$ is $p_j - p_k$. Let

$$z = \frac{(p_j - p_k) - \pi_0}{s_e(p_j - p_k)} \quad (20)$$

where $s_e(p_j - p_k)$ is the standard deviation of $p_j - p_k$ given as

$$s_e(p_j - p_k) = \sqrt{\text{Var}(p_j) + \text{Var}(p_k) - 2\text{Cov}(p_j, p_k)}$$

where

$$\begin{aligned} \text{Cov}(p_j, p_k) &= E[p_j p_k] - E[p_j]E[p_k] \\ &= \frac{E[t_j t_k]}{r^2} - \frac{E[t_j]E[t_k]}{r}, \text{ see equation (8)} \\ &= \frac{\sum_{i=1}^r \sum_{s=1}^r E[u_{ij} u_{sk}]}{r^2} - \pi_j \pi_k, \text{ see equation (7)} \end{aligned}$$

Now $u_{ij} u_{sk}$ assumes the value 1 if and only if u_{ij} and u_{sk} both assume the value 1 with probability $\pi_j \pi_k$.

Therefore,

$$\frac{\sum_{i=1}^r \sum_{s=1}^r E[u_{ij} u_{sk}]}{r^2} = \frac{r^2 (\pi_j \pi_k)}{r^2} = \pi_j \pi_k$$

PAIRWISE COMPARISON IN REPEATED MEASURES

Hence,

$$Cov(p_j, p_k) = \pi_j \pi_k - \pi_j \pi_k = 0$$

Therefore,

$$Var(p_j - p_k) = Var(p_j) + Var(p_k) \quad (21)$$

Under the null hypothesis of (18) the statistic z of (20)

$$z = \frac{(p_j - p_k) - \pi_o}{s_e(p_j - p_k)} = \frac{(p_j - p_k) - \pi_o}{\sqrt{Var(p_j) + Var(p_k)}}$$

which is the unit normal distribution. Hence,

$$\chi^2 = z^2 = \frac{((p_j - p_k) - \pi_o)^2}{Var(p_j) + Var(p_k)} \quad (22)$$

has approximately a chi-square distribution with 1 degree of freedom where p_j is given in (8) and

$$Var(p_j) = \frac{t_j(r - t_j)}{r^3}, \text{ see equation (9).}$$

Under the null hypothesis of no difference between the population or treatment proportions of success and overall estimate of π_j namely p_j is p given in (10). Hence the variance of p_j may be estimated as

$$Var(p_j) = \frac{p(1-p)}{r} = \frac{t(r(c-1)-t)}{r^3(c-1)^2} \quad (23)$$

Therefore, the test statistic for testing the null hypothesis, H_0 of Equation (18), is given in Equation (24) as:

$$\chi^2 = z^2 = \frac{\left(\left(\frac{t_j}{r} - \frac{t_k}{r}\right) - \pi_o\right)^2}{\text{Var}(p_j) + \text{Var}(p_k)} = \frac{\left(\left(\frac{t_j}{r} - \frac{t_k}{r}\right) - \pi_o\right)^2}{2\text{Var}(p)} \quad (24)$$

where under H_0 ,

$$\begin{aligned} \text{Var}(p_j) &= \text{Var}(p_k) = \text{Var}(p) \\ &= \frac{t(r(c-1)-t)}{r^3(c-1)^2}, \text{ see equation (23)} \end{aligned}$$

The test statistics of (22) may be expressed as

$$\chi^2 = \frac{r^3(c-1)^2 \left(\frac{t_j}{r} - \frac{t_k}{r}\right)}{t(r(c-1)-t)} - \pi_o \quad (25)$$

or equivalently in terms of the proportions in (8) and (10)

$$\chi^2 = \frac{r((p_j - p_k) - \pi_o)^2}{2p(1-p)} \quad (26)$$

which has approximately a chi-square distribution with 1 degree of freedom.

The test statistics of equation (26) may be used to test the null hypothesis that the proportion of positive responses or successes in the j^{th} treatment condition is higher than the corresponding proportion in the k^{th} treatment condition by at least some value, π_o . The statistics of equation (26) may be compared with an appropriately chosen critical value of the chi-squared distribution with degrees of freedom at a specified significance level α . However to keep the type 1 error small and control for erroneous conclusions, it is suggested that all comparisons be made against critical chi-square values with $c - 1$ degrees of freedom at a specified α level.

PAIRWISE COMPARISON IN REPEATED MEASURES

To test the null hypothesis of equation (19), that is that the positive response or success rate at the j^{th} experimental or treatment condition is greater than the corresponding success rate at the $(j - 1)^{\text{th}}$ experimental condition by at least some constant π_o , we use the test statistics

$$\chi^2 = \frac{(p_j - \pi_o)^2}{\text{Var}(p_j)}$$

Under H_o this becomes

$$\chi^2 = \frac{(p_j - \pi_o)^2}{\text{Var}(p_j)} = \frac{r^3 (c-1)^2 \left(\frac{t_j}{r} - \pi_o \right)^2}{t(r(c-1) - t)} = \frac{r(p - \pi_o)^2}{p(1-p)} \quad (27)$$

which has approximately a chi-square distribution with 1 degree of freedom. The test statistics of equation (27) may therefore similarly be used to test the null hypothesis that the proportion of positive responses or successes in the current, that is in the j^{th} treatment condition is at least equal to the corresponding proportion of positive responses in the immediately preceding, that is in the $(j - 1)^{\text{th}}$ treatment on experimental condition, although as suggested earlier all comparisons should preferably be made against critical chi-square values with $c - 1$ degrees of freedom.

If subjects or candidates are required to significantly improve their success rates early in the study or experiment, for example, at the second trial so that the success rate at the second trial is expected to be significantly greater than the success rate in the first trial then the null hypothesis of no difference between the success rates in these two trials must be rejected. This implies that for a given value of r and specified α level, the test statistics of equation (27) must be such that

$$\chi^2 = \frac{r^3 \left(\frac{t_2}{r} - 0 \right)^2}{t_2(r - t_2)} \geq \chi_{1-\alpha;1}^2,$$

and thus

$$\frac{p_2}{1-p_2} \geq \frac{\chi^2_{1-\alpha;1}}{r} \quad (28)$$

(28) is an estimate of the odds that a randomly selected candidate or subject performs significantly better in the second than in the first trial. In other words for a given number of subjects r and specified significance level α , the estimated probability p_2 that a randomly selected subject for some experiment or study, performs significantly better or significantly improves his performance at the second trial relative to the first trial must be such that

$$p_2 \geq \frac{\chi^2_{1-\alpha;1}}{r + \chi^2_{1-\alpha;1}} \quad (29)$$

Illustrative Examples

Example 1

If $r = 20$ and $\alpha = 0.05$ or 0.01 , then

$$p_2 \geq \frac{3.841}{23.841} = 0.161, \text{ or } p_2 \geq \frac{5.991}{25.991} = 0.23$$

Under these circumstances a randomly selected subject or candidate in the program of interest would have to earn at least 16.1 percent or 23.0 percent higher in the current or second trial relative to the immediately preceding or first trial to be able to continue with the experiment or program.

Example 2

Shown in [Table 3](#) are the grade-point-averages (GPAs) earned in each of the five years a random sample of 23 students who studied medical statistics at a certain university.

PAIRWISE COMPARISON IN REPEATED MEASURES

Table 3. Grade Point Averages (GPA) of a Random Sample of Students

Student	Year of Study				
	Year 1	Year 2	Year 3	Year 4	Year 5
1	2.06	2.05	2.29	2.67	1.00
2	1.38	2.08	2.42	2.76	1.00
3	3.54	3.71	3.88	3.57	5.00
4	1.33	2.17	2.33	1.86	2.22
5	2.02	2.37	2.29	2.60	1.00
6	3.08	3.30	3.36	3.73	1.00
7	1.21	2.30	2.70	2.00	2.44
8	1.35	2.25	2.06	2.00	2.44
9	1.88	1.82	3.64	3.39	2.00
10	2.06	3.14	2.44	3.00	1.00
11	1.85	2.50	2.51	2.80	1.00
12	1.94	1.39	1.83	1.06	1.00
13	2.91	3.39	2.91	2.13	4.00
14	4.16	2.17	1.57	1.21	1.25
15	1.50	1.90	1.50	1.79	3.50
16	1.54	2.85	2.73	1.00	1.50
17	1.96	2.21	2.57	2.00	1.00
18	1.24	2.29	1.04	1.74	2.00
19	1.26	2.67	1.20	1.18	1.50
20	1.49	2.21	2.57	1.71	1.00
21	1.55	2.50	1.86	1.21	2.00
22	2.46	2.39	3.56	3.50	1.00
23	1.33	2.16	1.25	2.12	1.43

The research interest is to determine whether or not students in the program progressively improved their academic performance. To answer this question apply (1) with the data of Table 2 to generate the corresponding coded data of 1's and 0's shown in Table 4.

Table 4. Patterns of 1's and 0's for Data in Table 3.

Student	Year of study				Total
	Year 2	Year 3	Year 4	Year 5	
1	0	1	1	0	--
2	1	1	1	0	--
3	1	1	0	1	--
4	1	1	0	1	--
5	1	0	1	0	--
6	1	1	1	0	--
7	1	1	0	1	--
8	1	0	1	0	--
9	0	1	0	0	--
10	1	0	1	0	--
11	1	1	1	0	--
12	0	1	0	0	--
13	1	0	0	1	--
14	1	0	0	1	--
15	1	0	1	1	--
16	1	0	0	1	--
17	1	1	0	0	--
18	1	0	1	1	--
19	1	0	0	1	--
20	1	1	0	0	--
21	1	1	0	1	--
22	0	1	0	0	--
23	1	0	1	0	--
No of 1's (t_j)	19	12	10	10	51
No of 0's ($r-t_j$)	4	11	13	13	41
Total r	23	23	23	23	92
Proportion of 1's (p_j)	0.826	0.522	0.435	0.435	0.554

Now using the proportion of 1's or successes of Table 3 in (4), we have

$$\begin{aligned}
 \chi^2 &= \frac{(23)(0.272)^2 + (-0.032)^2 + (-0.119)^2 + (-0.119)^2}{(0.554)(0.446)} \\
 &= \frac{(23)(0.074 + 0.001 + 0.014 + 0.014)}{0.247} \\
 &= \frac{(23)(0.103)}{0.247} = \frac{2.3695}{0.247} = 9.591
 \end{aligned}$$

PAIRWISE COMPARISON IN REPEATED MEASURES

With $c - 1 = 5 - 1 = 4$ degrees of freedom it is statistically significant at $\alpha = 0.05$ ($\chi^2_{0.95,4} = 9.488$). This means that students' performance seems to differ significantly from year to year. In fact the results of Table 3 suggest that students' success rates declined progressively during the years of study. Further interest may now be in comparing some of the years in terms of success rates achieved to determine which ones may be responsible for the rejection of H_o . For example, one may wish to compare year 5 with year 2 to determine whether there is any significant difference in the relative success rates for the two years. That is interest may be in testing the null hypothesis of (18) with $\pi_o \geq 0$.

Again using the data of Table 3 with $p_5 = 0.435$ and $p_2 = 0.826$ in (26)

$$\chi^2 = \frac{(23)(0.435 - 0.826)^2}{(2)(0.554)(0.446)} = \frac{3.516}{0.4942} = 7.117$$

With $c - 1 = 5 - 1 = 4$ degrees of freedom it is not statistically significant at $\alpha = 0.05$.

Similar comparisons can also be made for other pairs of years. Pairwise comparisons for other years (e.g., year 2 versus year 1, year 3 versus year 2) may be conducted to ascertain whether the relative success rates for these pairs of years are statistically different from zero. Thus for year 3 versus year 2, from Table 3 that $p_3 = 0.522$. From (27) with $\pi_o = 0$

$$\chi^2 = \frac{(23)(0.522)^2}{(0.554)(0.446)} = \frac{6.267}{0.247} = 25.372$$

which with 4 degrees of freedom is statistically significant at $\alpha = 0.05$. Similarly for year 2 relative to year 1, $p_2 = 0.826$,

$$\chi^2 = \frac{(23)(0.826)^2}{(0.554)(0.446)} = \frac{15.692}{0.247} = 63.530$$

and with 4 degrees of freedom is also statistically significant at $\alpha = 0.05$.

These results indicate that students significantly improved their performances during their second and third years of study relative to the immediately preceding years. Note that if it is required that students must achieve some critical minimum

score at the end of the second year relative to the first year, then if $\alpha=0.05$, the required minimum score is from equation (29).

$$p_2 \geq \frac{3.841}{23+3.841} = \frac{3.841}{26.841} = 0.143, \text{ or } 14.3 \text{ percent}$$

which is much less than the average success rate of 82.6 percent achieved in year 2 relative to year 1. Note also that for year 3 and year 2,

$$\chi^2 = \frac{23(0.522-0.826)^2}{2(0.544)(0.446)} = \frac{2.126}{0.486} = 4.384$$

which with 4 degrees of freedom is not statistically significant at $\alpha = 0.05$. This implies that there is no statistical difference between the improvement rates of students at their junior year relative to their sophomore year and the improvement rates at the sophomore year relative to the freshman years. In other words the students may have improved their performances equally well in their third year relative to their second year as in the second relative the first year.

Other values of q_{jk} are similarly calculated and the results are presented in Table 5.

Table 5. Values of q_{jk} (Under $H_0: \pi_o = 0$)

<i>j</i>	<i>k</i>				Marginal
	2 (0.826)	3 (0.522)	4 (0.435)	5 (0.435)	
2 (0.826)	--	--	--	--	63.530
3 (0.522)	4.302	--	--	--	23.372
4 (0.435)	7.117	0.352	--	--	17.619
5 (0.435)	7.117	0.352	0.000	--	17.619

Note: *j* = Current Trial Relative Proportions; *k* = Immediate Past Trial Relative Proportions

All chi-square values in Table 5 which are at least equal to 9.488, the critical chi-square value with 4 degrees of freedom at $\alpha = 0.05$ are statistically significant. Therefore, all the years are here responsible for the observed significant difference in the relative success rates of students during their study period.

In order to illustrate the efficiency of this proposed method, Friedman Two-way Analysis Of Variance is applied to the above data. The observations within

PAIRWISE COMPARISON IN REPEATED MEASURES

each row (r) or blocks are ranked from the largest to the smallest with the rank of '1' assigned to the largest value and the rank of '2' assigned to the next largest value and so on, until the rank c is assigned to the smallest value, where c is the number of treatments. The test statistics for ($r \geq 10, c \geq 4$) are given by

$$\chi^2 = \frac{12}{rc(c+1)} \sum_{j=1}^c R_j^2 - 3r(c+1)$$

Table 6. An illustration of the Friedman Two-way Analysis of Variance by Ranks.

Student	Year 1	Year 2	Year 3	Year 4	Year 5
1	2.06(3)	2.05(4)	2.29(2)	2.67(1)	1.00(5)
2	1.38(5)	2.08(3)	2.42(2)	2.76(1)	2.00(4)
3	3.54(4)	3.71(3)	3.88(2)	3.51(5)	5.00(1)
4	1.33(5)	2.17(3)	2.33(1)	1.86(4)	2.22(2)
5	2.02(4)	2.37(2)	2.29(3)	2.60(1)	1.00(5)
6	3.08(4)	3.30(3)	3.36(2)	2.73(1)	1.00(5)
7	1.21(5)	2.30(3)	2.70(1)	2.00(4)	2.44(2)
8	1.35(5)	2.25(2)	2.06(3)	2.43(1)	2.00(4)
9	1.88(4)	1.82(5)	3.64(1)	3.39(2)	2.00(3)
10	2.06(4)	3.14(1)	2.44(3)	3.00(2)	1.00(5)
11	1.85(4)	2.50(3)	2.51(2)	2.80(1)	1.00(5)
12	1.94(1)	1.39(3)	1.83(2)	1.06(4)	1.00(5)
13	2.91(3.5)	3.39(2)	2.91(3.5)	2.13(5)	4.00(1)
14	1.16(5)	2.17(1)	1.57(2)	1.21(4)	1.25(3)
15	1.50(4.5)	1.90(2)	1.50(4.5)	1.79(3)	3.50(1)
16	1.54(3)	2.85(1)	2.73(2)	1.00(5)	1.50(4)
17	1.96(4)	2.21(2)	2.57(1)	2.00(3)	1.00(5)
18	1.24(4)	2.29(1)	1.04(5)	1.74(3)	2.00(2)
19	1.26(3)	2.67(1)	1.20(4)	1.18(5)	1.50(2)
20	1.49(4)	2.21(2)	2.57(1)	1.71(3)	1.00(5)
21	1.55(5)	2.50(1)	1.86(3)	1.21(5)	2.00(2)
22	2.46(3)	2.39(4)	3.56(1)	3.50(2)	1.00(5)
23	1.33(4)	2.16(1)	1.25(5)	2.12(2)	1.43(3)
Rank Sum (R_j)	90	53	56	67	79

Now using

$$\begin{aligned}\chi^2 &= \frac{12}{rc(c+1)} \sum_{i=1}^c R_j^2 - 3r(c+1) \\ \chi^2 &= \frac{12}{23(5)(5+1)} (90^2 + 53^2 + 56^2 + 67^2 + 79^2) - 3(23)(5+1) \\ \chi^2 &= \frac{12}{rc(c+1)} \sum_{i=1}^c R_j^2 - 3r(c+1) \\ \chi^2 &= 16.76\end{aligned}$$

which is also significant at $\alpha = 0.05$, because $\chi^2 > \chi_{1-\alpha; c-1}^2$, $16.76 > 9.488$, and $\chi_{1-\alpha; c-1}^2 = \chi_{1-0.05; 5-1}^2 = \chi_{0.95; 4}^2 = 9.488$. The results are also significant like the proposed methods.

Conclusion

The proposed method of pairwise comparisons in repeated measures is suitable when interest is not only on testing whether the null hypothesis of no difference is rejected or accepted, but if the null hypothesis is rejected, which individual subjects or their combinations actually contributed to the rejection of the null hypothesis. It is at least as efficient and powerful as competitors.

References

- Conover, W. J. (1980). *Practical Nonparametric Statistic*. Second Ed. New York: John Wiley & Sons, Inc.
- Davidian, M & Giltinan, D. M. (1995). *Nonlinear Models for Repeated Measurement Data*. Boca Raton, FL: Chapman & Hall/CRC.
- Dean, A. & Voss, D. (1999). *Design and Analysis of Experiments*. New York: Springer-Verlag.
- Fitzmaurice, G., Davidian, M., Verbeke, G., & Molenberghs, G., Eds. (2008). *Longitudinal Data Analysis*. Boca Raton, FL: Chapman & Hall/CRC.
- Friedman, M. (1937). The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. *Journal of the American Statistical Association*, 32(200), 675-701.

PAIRWISE COMPARISON IN REPEATED MEASURES

- Girden, E. R. (1992). *ANOVA Repeated Measures*. New York: Sage Publications.
- Hollander, M. & Wolfe, D. A. (1979). *Non-parametric Statistical Methods*, New York: John Wiley & Sons, Inc.
- Jones, B. & Kenward, M. G. (2003). *Design and Analysis of Cross-Over Trials*, (2nd Ed.). London: Chapman and Hall.
- Kempthorne, O. (1979). Sampling Inference, Experimental Inference and Observation Inference. *Sankhya*, 40, 115-145.
- Oyeka, I. C. A. (1996). *Introduction to applied Statistical Methods*. Enugu, Nigeria: Nobern Allocation Publishing Company.
- Oyeka, I. C. A. (1990). A Non-Parametric Method for the Analysis of Repeated Responses. *STATISTICA*, 50(1), 83-90.
- Oyeka, I. C. A. (1986). Family Planning Among Nigerian Postsecondary Female Students. *Studies in Family Planning*, 17(3), 146-152.
- Page, E. B. (1963). Ordered Hypothesis for Multiple Treatments: A Significant Test for Linear Ranks. *Journal of American Statistical Association*, 58(301), 216-230.
- Pan, J-X. & Fang, K-T. (2002). Growth Curve models and Statistical diagnostics. *Springer Series in Statistics*. New York: Springer-Verlag.
- Prentice, R. L. (1978). Linear Rank Tests with Right Censored Data. *BIOMETRIKA*, 65, 167-179.
- Sen, P. K. (1968). Robustness of Some Nonparametric Procedures in Linear Models. *The Annals of Mathematical Statistics*, 39(6), 1913-1922.
- Sen, P. K. & Puri, M. L. (1967). On the Theory of Rank Order Tests for Location in the Multivariate One Sample Problem. *The Annals of Mathematical Statistics*, 38(4), 1216-1228.
- Shanghnessy, J. J. (2006). *Research methods in Psychology*. New York: McGraw-Hill.
- Timm, N. H. (2002). Applied Multivariate Analysis. *Springer Texts in Statistics*. New York: Springer-Verlag.
- Vonesh, E. F. & Chinchilli, V. G. (1997). *Linear and Nonlinear Models for Analysis of Repeated Measurements*. London: Chapman and Hall.

A Bivariate Distribution with Conditional Gamma and its Multivariate Form

Sumen Sen

Old Dominion University
Norfolk, VA

Rajan Lamichhane

Texas A&M University -
Kingsville, TX

Norou Diawara

Old Dominion University
Norfolk, VA

A bivariate distribution whose marginal are gamma and beta prime distribution is introduced. The distribution is derived and the generation of such bivariate sample is shown. Extension of the results are given in the multivariate case under a joint independent component analysis method. Simulated applications are given and they show consistency of our approach. Estimation procedures for the bivariate case are provided.

Keywords: Gamma distribution, Gamma function, Beta function, Beta distribution, generalized Beta prime distribution, incomplete gamma function

Introduction

The gamma and beta distributions are the two most commonly used distribution when it comes to analyzing skewed data. Since Kibble (1941), the bivariate gamma has gained considerable attention. The multivariate form of the gamma has been proposed in Johnson et al. (1997) and by many other authors, but there is no unifying formulation. Even in the multivariate exponential family of distributions, there is no known multivariate gamma (Joe, 1997). The simplest of the multivariate cases, the bivariate gamma distribution, is still raising debates, and has been proposed in Balakrishnan and Lai (2009). The marginal densities of the bivariate gamma can sometimes belong to other class of distributions. A modified version of Nadarajah (2009) bivariate distribution with Gamma and Beta marginals is considered, and a conditional component to the modeling is brought into account. Kotz et al (2004) proposed a bivariate gamma exponential distribution with gamma and Pareto distribution as marginals. In this article, a bivariate gamma distribution

Sumen Sen is a Graduate Student Assistant in the Department of Mathematics and Statistics. Email him at sxsen002@odu.edu. Dr. Lamichhane is an Assistant Professor in the Department of Mathematics. Email him at rajan.lamichhane@tamuk.edu. Dr. Diawara is an Associate Professor in the Department of Mathematics and Statistics. Email him at ndiawara@odu.edu.

with gamma and beta prime as marginal distributions is defined. By including the dependence structure, more flexibility is added. Consider two random variables X , identified as the common measure, and Y related to X , and assuming that X is a gamma random variable with parameters α and β and the distribution of $Y|X$ is a gamma random variable with parameters a and X . The first section following this introduction shows the bivariate distribution with the conditional gamma. In the next section, ‘Properties,’ the main properties of the bivariate conditional gamma distribution are given. Extension to the multivariate setting is given in the next section, followed by a development of computational aspects in the inference. The calculations in this paper involve several special functions, including the incomplete gamma function defined by

$$\gamma(a, x) = \frac{1}{\Gamma(a)} \int_0^x t^{a-1} e^{-t} dt,$$

and the complementary gamma function defined as

$$\Gamma(a, x) = \int_x^\infty t^{a-1} e^{-t} dt, \text{ with } \Gamma(a) = \Gamma(a, 0).$$

Also, the beta function is defined as

$$B(a, b) = \int_0^1 t^{a-1} (1-t)^{b-1} dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)},$$

for a , b , and x positive real values. For $x \in [0, 1]$, $\alpha > 0$ and $\beta > 0$, the beta distribution can be defined as

$$f(x) = \frac{x^{\alpha-1} (1-x)^{\beta-1}}{B(\alpha, \beta)}$$

Model Building and Density functions

Let X be a gamma rv's with shape and rate parameters denoted by α and β , respectively. The probability density function (pdf) of X is given by

$$f_x(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, x > 0 \quad (1)$$

where $\alpha > 0$ and $\beta > 0$. Many authors have developed structural models with the underlying gamma distribution. Consider another random variable Y such that the distribution of the random variable Y given a realization of X at x is a gamma with the parameters a and x . That is the density of $Y|X$ is given by

$$f_{Y|X}(y|x) = \frac{x^a}{\Gamma(a)} y^{a-1} e^{-xy}, y > 0 \quad (2)$$

where $a > 0$ and $x > 0$ are the shape and rate parameters respectively. So the joint density of the random variables defined above is given by the expression below

$$\begin{aligned} f_{X,Y}(x,y) &= f_{Y|X=x}(y|X=x) * f_X(x) \\ &= \frac{\beta^\alpha x^{\alpha+a-1} y^{a-1}}{\Gamma(a)\Gamma(\alpha)} e^{-\beta x} e^{-xy}, x > 0 \text{ and } y > 0, \end{aligned} \quad (3)$$

with parameters $\alpha > 0$, $\beta > 0$ and $a > 0$. Equation (3) integrates to 1, so this is a legitimate distribution. Figure 1 shows the plot of the joint distribution defined in Equation (3) for different values of α , β and a .

Thus the cumulative distribution of the random variable X and Y is

$$F_{X,Y}(x,y) = \frac{\gamma(\alpha, \beta x) \gamma(a, xy)}{\Gamma(a)\Gamma(\alpha)}, x > 0 \text{ and } y > 0. \quad (4)$$

A BIVARIATE DISTRIBUTION WITH CONDITIONAL GAMMA

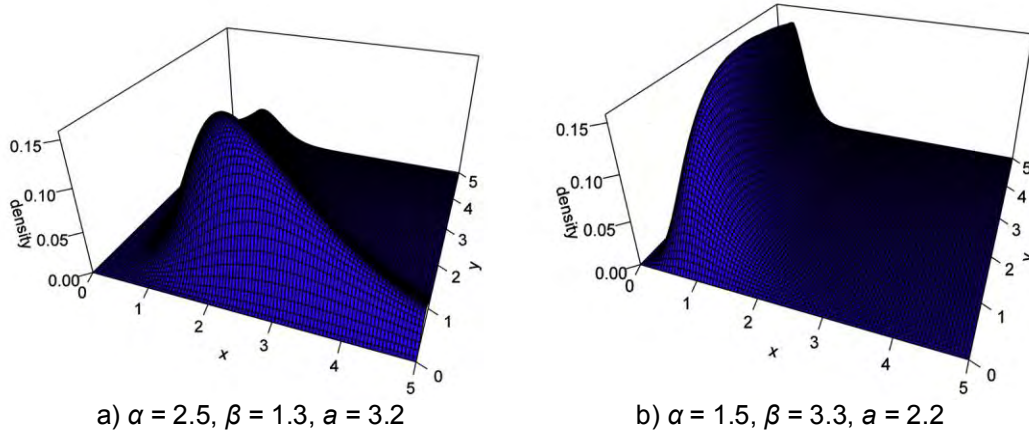


Figure 1. Joint Probability Density Function of (X, Y)

Properties

The main properties of the distribution as defined in (3), such as the marginal densities, their moments, their product products and covariance, are derived here.

Marginal Density and Moments:

The marginal density of X is given by (1). Marginal density of Y is given by the theorem below.

Theorem 1: If the joint density of (X, Y) is given in (3), then the marginal density of Y is given by

$$f_y(y) = \frac{1}{B(a, \alpha)} \frac{\beta^\alpha y^{a-1}}{(y + \beta)^{a+\alpha}}, \quad y > 0, a > 0 \text{ and } \alpha > 0 \quad (5)$$

Proof. The marginal density of Y is given by

$$\begin{aligned} f_y(y) &= \frac{\beta^\alpha y^{a-1}}{\Gamma(\alpha)\Gamma(a)} \int_0^\infty x^{a+\alpha-1} e^{-(\beta x+xy)} dx \\ &= \frac{\beta^\alpha y^{a-1}}{\Gamma(\alpha)\Gamma(a)} \int_0^\infty \frac{z^{a+\alpha-1} e^{-z}}{(y+\beta)^{a+\alpha}} dz, \text{ with } z = x(y+\beta) \\ &= \frac{\Gamma(a+\alpha)}{\Gamma(a)\Gamma(\alpha)} \frac{\beta^\alpha y^{a-1}}{(y+\beta)^{a+\alpha}} \end{aligned}$$

Probability density function of Y is a special form of Generalized Beta prime density with shape parameter 1 and scale parameter β . Figure 2 describes its pdf for different values of α . Probability density of generalized beta prime distribution with scale p and shape q is given by

$$f(y; \alpha, \beta, p, q) = \frac{p \left(\frac{x}{q} \right)^{\alpha p - 1} \left(1 + \left(\frac{x}{q} \right)^p \right)^{-\alpha - \beta}}{q B(\alpha, \beta)} \quad (6)$$

Let T be a random variable such that $T \sim \text{Beta}(a, \alpha)$. Then $Y = \frac{\beta T}{1-T}$, has density given by (5).

Theorem 2: Let Y be a random variable with a pdf given in (5). The m^{th} moment of the random variable Y exists only if $\alpha > m$.

Proof: From the previous theorem it can be seen that if $T: \text{Beta}(a, \alpha)$ and $Y = \frac{\beta t}{1-t}$, then the density of Y will be same as defined in (5). And the m^{th} moment of Y is

A BIVARIATE DISTRIBUTION WITH CONDITIONAL GAMMA

$$\begin{aligned}
 E[Y^m] &= E\left[\frac{\beta t}{1-t}\right]^m \\
 &= \int_0^1 \frac{\beta^m t^m}{(1-t)^m} \frac{t^{a-1} (1-t)^{\alpha-1}}{\text{Beta}(a, \alpha)} \\
 &= \frac{\beta^m}{\text{Beta}(a, \alpha)} \int_0^1 t^{a+m-1} (1-t)^{\alpha-m-1} \\
 &= \frac{\beta^m}{\text{Beta}(a, \alpha)} \text{Beta}(a+m, \alpha-m), \text{ provided } \alpha > m \\
 &\quad \text{with } \text{Beta}(a, \alpha) = \frac{\Gamma(a)\Gamma(\alpha)}{\Gamma(a+\alpha)}
 \end{aligned}$$

The choice of $\alpha > 2$ is made so that $E[Y]$ and $\text{Var}[Y]$ will both exist.

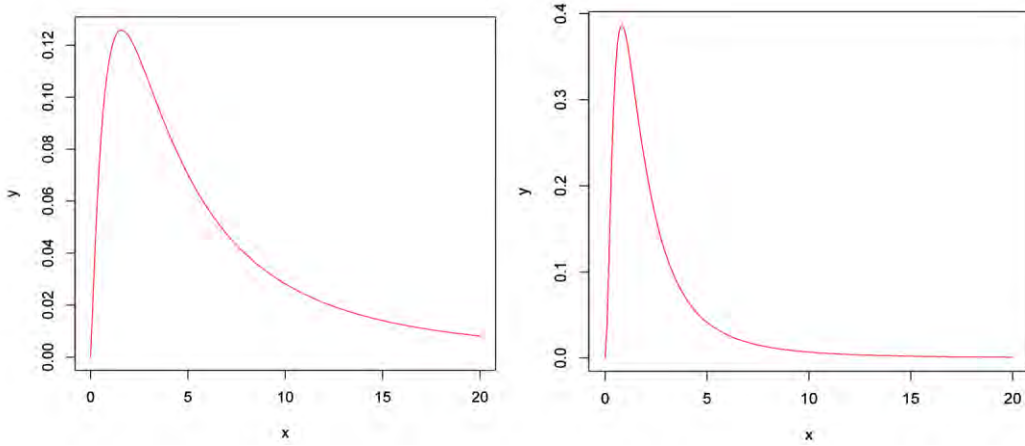


Figure 2. The Probability Density Function of Y as defined in (5)

Product Moments

Theorem 3: The product moment of the random variables (X, Y) associated with the pdf defined in (3) can be expressed as

$$E(X^m Y^n) = \frac{\beta^{n-m} \Gamma(a+n) \Gamma(m-n+\alpha)}{\Gamma(a)} \text{ for } a > 0, m > 0, n > 0 \text{ and } m + \alpha > n$$

Proof: For $m > 0$ and $n > 0$ one can write

$$\begin{aligned} E(X^m Y^n) &= \int_0^\infty \int_0^\infty x^m y^n \frac{\beta^\alpha x^{a+\alpha-1} y^{a-1} e^{-\beta x - xy}}{\Gamma(\alpha) \Gamma(a)} dx dy \\ &= \int_0^\infty \frac{\beta^\alpha e^{-\beta x} x^{a+m+\alpha-1}}{\Gamma(\alpha) \Gamma(a)} \int_0^\infty y^{a+n-1} e^{(-xy)} dy dx \\ &= \int_0^\infty \frac{\beta^\alpha x^{a+m+\alpha-1} e^{-\beta x}}{\Gamma(\alpha) \Gamma(a)} \frac{\Gamma(a+n)}{x^{a+n}} dx \\ &= \frac{\beta^\alpha \Gamma(a+n)}{\Gamma(\alpha) \Gamma(a)} \int_0^\infty e^{-\beta x} x^{m-n+\alpha-1} dx \\ &= \frac{\beta^{n-m} \Gamma(a+n) \Gamma(m-n+\alpha)}{\Gamma(\alpha) \Gamma(a)} \end{aligned}$$

provided the integrals exist. Now for the m^{th} product moment by choosing $n = m$ in the above expression, one can write the product moment as

$$E(X^m Y^n) = \frac{\Gamma(a+m)}{\Gamma(a)}$$

Note that the product moment depends only on a .

Covariance Matrix

With the density of X and Y as given in Equations (1) and (5), respectively, the variance-covariance matrix of X and Y is given by the following theorem

Theorem 4: Denote the Variance-Covariance matrix of X and Y by $Cov(X, Y)$, then

$$Cov(X, Y) = \begin{pmatrix} \frac{\alpha}{\beta^2} & \frac{-a}{\alpha-1} \\ \frac{-a}{\alpha-1} & \frac{a\beta^2(a+\alpha-1)}{(\alpha-1)^2(\alpha-2)} \end{pmatrix} \quad (7)$$

Proof: Using Theorem 2, the variance and expectation of the random variable Y can be computed

$$\begin{aligned} Var(Y) &= E(Y^2) - (E(Y))^2 \\ &= \frac{a\beta^2(a+1)}{(\alpha-1)(\alpha-2)} - \left[\frac{a\beta}{(\alpha-1)} \right]^2 \\ &= \frac{a\beta^2(a+\alpha-1)}{(\alpha-1)^2(\alpha-2)} \end{aligned} \quad (8)$$

Equation (2) implies that the distribution of X is a Gamma distribution with shape α and rate β . So variance of X is given by

$$Var(X) = \frac{\alpha}{\beta^2} \quad (9)$$

Now the covariance between X and Y can be written as

$$\begin{aligned}
 Cov(X, Y) &= E(XY) - E(X)E(Y) \\
 &= E(XE(Y|X)) - E(X)E(Y) \\
 &= E\left(X \frac{a}{X}\right) - E(X)E(Y) \text{ as } Y|X : Gamma(a, x) \quad (10) \\
 &= a - \frac{\alpha}{\beta} \frac{a\beta}{(\alpha-1)} \\
 &= -\frac{a}{(\alpha-1)}
 \end{aligned}$$

Using the Equations (8), (9) and (10) the result follows. Note that the covariance between X and Y exists only when $\alpha \neq 1$, and is positive when $\alpha < 1$. Variance of Y only exists when $\alpha > 2$.

Multivariate Extension Case

Consider the multivariate case of the model: take $n + 1$ random variables as follows:

$$\begin{aligned}
 X_0 &: Gamma(\alpha, \beta) \\
 X_1 | X_0 &: Gamma(a_1, b_1 x_0) \\
 X_2 | X_0 &: Gamma(a_2, b_2 x_0) \\
 &\vdots \\
 X_n | X_0 &: Gamma(a_n, b_n x_0)
 \end{aligned}$$

where $X_i | X_0$ and $X_j | X_0$ are independent components for $i \neq j$ and $(i, j) \in \{1, 2, \dots, n\}$. Then using the same argument as in ‘[Properties](#),’ the joint independent component model is built and the marginal density function for each random variable X_i is derived. In general, the density function of X_i is given by

$$f(x_i) = \frac{\Gamma(\alpha + a_i)}{\Gamma(a_i)\Gamma(\alpha)} \frac{b_i^{a_i} \beta^\alpha x_i^{a_i-1}}{(\beta + b_i x_i)^{\alpha+a_i}}, \text{ for } i = 1, 2, \dots, n \quad (11)$$

Using the independence assumption of the above model, the joint density of X_0, X_1, \dots, X_n is then derived. The derived joint density will be of the form

A BIVARIATE DISTRIBUTION WITH CONDITIONAL GAMMA

$$f(x_0, x_1, \dots, x_n) = f(x_0) \prod_{i=1}^{i=n} f(x_i | x_0)$$

The density of the joint distribution (X_1, X_2, \dots, X_n) and its variance covariance expression are derived next.

For the density of (X_1, X_2, \dots, X_n) , the integration of the joint density with respect to the variable X_0 is needed.

Density Function

To derive the density function, the integral below is computed

$$f(x_1, \dots, x_n) = \int_0^\infty f(x_0) \prod_{i=1}^{i=n} f(x_i | x_0) dx_0 \quad (12)$$

And solving the integral in (12), the joint density is as follows

$$f(x_1, \dots, x_n) = \frac{\Gamma\left(\alpha + \sum_{i=1}^n a_i\right)}{\Gamma(\alpha) \sum_{i=1}^n \Gamma(a_i)} \frac{\beta^\alpha \prod_{i=1}^n (b_i^{a_i} x_i^{a_i-1})}{\left\{ \beta + \sum_{i=1}^n (b_i x_i) \right\}^{\sum_{i=1}^n a_i + \alpha}} \quad (13)$$

where $x_i > 0$, $a_i > 0$ for all $i = 1, 2, \dots, n$, and $\alpha > 0$, $\beta > 0$. In the distribution obtained from (13), if the choices of $\beta = 1$ and $b_i = 1$ for all $i = 1, 2, \dots, n$ are made, then the inverted Dirichlet distribution is obtained. The application of this distribution can be found in many places in the literature. Taio and Cuttman (1965) introduced this type of distribution and discussed about their applications.

Covariance

The covariance between X_i and X_j for $i \neq j$ is derived in Theorem 5.

Theorem 5: If the random variables X_1, X_2, \dots, X_n have the density function defined in (13), then the covariance between X_i and X_j for $i \neq j$ is given by the expression below

$$\text{Cov}(X_i, X_j) = \frac{a_i a_j \beta^2}{b_i b_j (\alpha - 1)^2 (\alpha - 2)} \text{ for } i \neq j, b_i \text{ can be equal to } b_j$$

Proof: Using the same arguments in [Theorem 2](#), the m^{th} moments of X_i are derived. Based on the density of X_i defined by (11)

$$E(X_i^m) = \frac{\Gamma(a_i + m) \Gamma(\alpha - m)}{\Gamma(a_i) \Gamma(\alpha)} \frac{\beta^m}{b_i^m} \text{ with } \alpha > m \quad (14)$$

From (13) this useful identity is obtained

$$\int_0^\infty \cdots \int_0^\infty \frac{\prod_{i=1}^n x_i^{a_i-1}}{\left(\beta + \sum_{i=1}^n b_i x_i \right)^{\alpha + \sum_{i=1}^n a_i}} = \frac{\Gamma(\alpha) \prod_{i=1}^n \Gamma(a_i)}{\beta^\alpha \sum_{i=1}^n b_i^{a_i} \Gamma\left(\alpha + \sum_{i=1}^n a_i\right)} \quad (15)$$

Using the identity in (15), the $(m_1, m_2, \dots, m_n)^{\text{th}}$ mixed moment is given as

$$E(X_i^{m_1} \dots X_i^{m_n}) = \frac{\Gamma\left(\alpha - \sum_{i=1}^n m_i\right)}{\Gamma(\alpha)} \prod_{i=1}^n \frac{\Gamma(a_i + m_i)}{b_i^{m_i} \Gamma(a_i)} \quad (16)$$

provided $\alpha > \sum_{i=1}^n m_i$. In particular, the covariances between X_i and X_j , for $i = 1, 2, \dots, n$, is as

$$\text{Cov}(X_i, X_j) = \frac{a_i a_j \beta^2}{b_i b_j (\alpha - 1)^2 (\alpha - 2)}$$

Note that the covariance between X_0 and X_i for $i = 1, 2, \dots, n$ is also derived as

$$\begin{aligned}
 \text{Cov}(X_0, X_i) &= E(X_0 X_i) - E(X_0)E(X_i) \\
 &= E(X_0 E(X_i | X_0)) - E(X_0)E(X_i) \\
 &= E\left(X_0 \frac{a_i}{X_0 b_i}\right) - E(X_0)E(X_i) \text{ as } X_i | X_0 : \text{Gamma}(a_i, b_i x) \quad (17) \\
 &= \frac{a_i}{X_0 b_i} - \frac{\alpha}{\beta} \frac{\beta \Gamma(a_i + 1) \Gamma(\alpha - 1)}{b_i \Gamma(a_i) \Gamma(\alpha)} \\
 &= -\frac{a_i}{b_i(\alpha - 1)}
 \end{aligned}$$

Bivariate cases will reduce to Equation (10).

Likelihood and Estimation for Bivariate Case

In this section, the maximum likelihood estimation process and Fisher information matrix for the bivariate model are introduced. Statistical analysis software (SAS) is used to generate data and R is used to get the maximum likelihood estimates (MLEs).

Log likelihood

Let (x_i, y_i) , for $i = 1, 2, \dots, n$, be a sample of size n from the bivariate gamma distribution as defined in Equation (3). Then, the log likelihood function is

$$\begin{aligned}
 L(x, y; \alpha, \beta, a) &= n\alpha \log(\beta) + (\alpha - 1) \sum_{i=1}^n \log(x_i) - n\beta \bar{x} - n \log[\Gamma(\alpha)] \\
 &\quad + a \sum_{i=1}^n x_i + (\alpha - 1) \sum_{i=1}^n \log(y_i) - \sum_{i=1}^n x_i y_i - n \log \Gamma(a) \quad (18)
 \end{aligned}$$

The first order derivatives of the log likelihood with respect to the three parameters are

$$\frac{\partial \text{Log}(\alpha, \beta, a)}{\partial \alpha} = n \log(\beta) + \sum_{i=0}^n \log(x_i) - n\psi(\alpha) \quad (19)$$

$$\frac{\partial \text{Log}(\alpha, \beta, a)}{\partial \beta} = \frac{n\alpha}{\beta} - n\bar{x} \quad (20)$$

$$\frac{\partial \text{Log}(\alpha, \beta, a)}{\partial a} = n\bar{x} + \sum_{i=0}^n \log(x_i) - n\psi(a) \quad (21)$$

where $\psi(x) = \frac{d}{dx} \ln(\Gamma(x))$ is the *Digamma* function.

Solving above Equations (19-21) simultaneously, the MLEs of the parameters can be formulated. As the MLEs are not in a closed form, an R code is developed to get the estimates.

Fisher Information Matrix

The Fisher information matrix g is given by the expectation of the covariance of partial derivatives of the log likelihood function. Let $(\theta_1, \theta_2, \theta_3) = (\alpha, \beta, a)$; then the components of the Fisher information matrix are given by

$$g_{ij} = \int_0^\infty \int_0^\infty \frac{\partial^2 \log f(x, y, \alpha, \beta, a)}{\partial \theta_i \partial \theta_j}$$

Hence, $g = \begin{pmatrix} \frac{\partial^2 \log \Gamma(\alpha)}{\partial \alpha} & \frac{-1}{\beta} & 0 \\ \frac{-1}{\beta} & \frac{\alpha}{\beta^2} & 0 \\ 0 & 0 & \frac{\partial^2 \log \Gamma(a)}{\partial a} \end{pmatrix} \quad (22)$

Inverting the fisher information matrix, the asymptotic standard errors of the maximum-likelihood estimates can be obtained.

Example Using Simulated Data

A number of simulations are performed to evaluate the statistical properties and the estimation are computed using *maximum likelihood method*. Because of the complexity of the target density and of the likelihood, there is no closed form of the estimators. Effective sample sizes will be directly impacting the estimates. R

A BIVARIATE DISTRIBUTION WITH CONDITIONAL GAMMA

program is used to do the optimization, but *SAS 9.3* version is used to simulate data with samples of sizes $n = 200$ and 25 .

Accordingly, for each set of parameters and sample size, X_0 is simulated from a gamma distribution with parameters α and β . Then, for each X_0 , generate X_1 based on X_0 according to Equation (3) with the same value of a .

The simulation results presented under the table give the estimates of the parameters. Figure 3 gives the plot of log likelihood and shows the uniqueness of the solution estimate for each parameter at sample size 200.

The results show that the larger the sample size, the more accurate the estimates are. A plot of the estimates versus sample size is given in Figure 4.

Simulation results

Table 1. Estimation of parameters for different sample sizes

Actual Values	Estimates (SE)	
	$n = 200$	$n = 25$
$\alpha = 2.5$	$\hat{\alpha} = 2.326 (0.819)$	$\hat{\alpha} = 2.001 (1.175)$
$\beta = 1.3$	$\hat{\beta} = 1.206 (0.474)$	$\hat{\beta} = 1.061 (0.708)$
$a = 3.2$	$\hat{a} = 3.229 (0.442)$	$\hat{a} = 3.075 (0.772)$
$\alpha = 6.3$	$\hat{\alpha} = 6.816 (2.503)$	$\hat{\alpha} = 5.287 (3.243)$
$\beta = 2.1$	$\hat{\beta} = 2.225 (0.848)$	$\hat{\beta} = 1.757 (1.131)$
$a = 1.2$	$\hat{a} = 1.169 (0.232)$	$\hat{a} = 1.280 (0.415)$

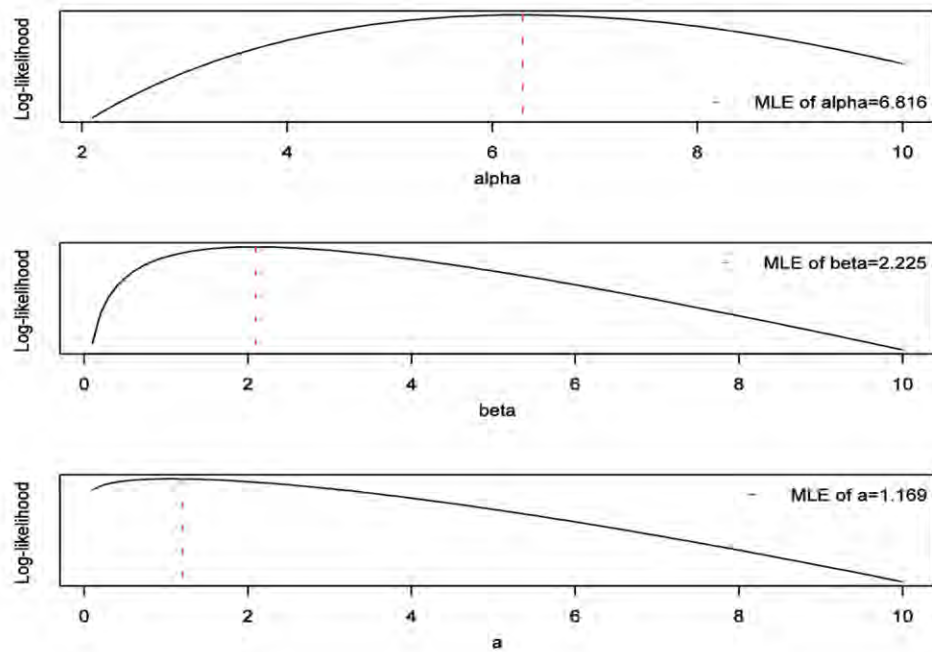


Figure 3. MLE estimates of parameters for a sample of 200

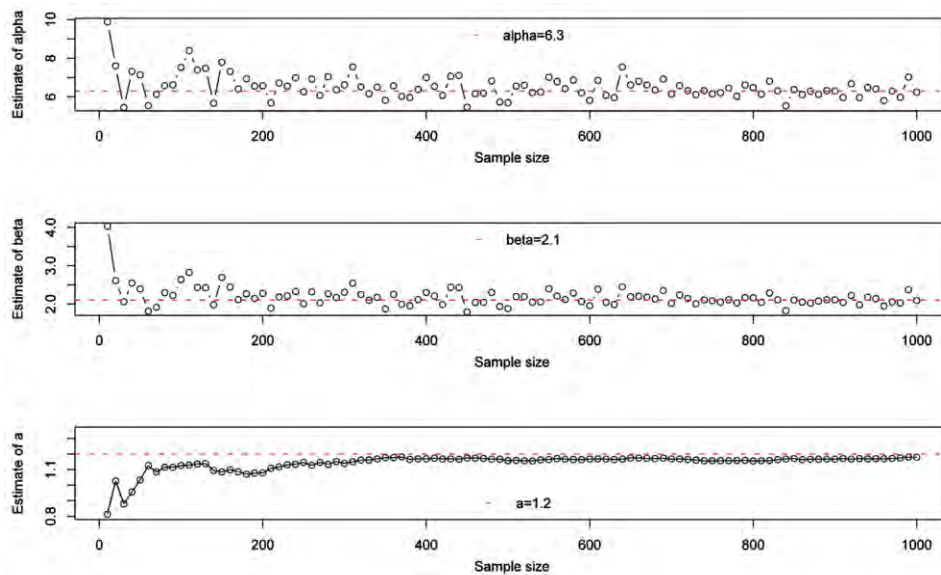


Figure 4. Parameter estimation for increasing sample size

Conclusion

In this paper, a bivariate conditional gamma and its multivariate form are proposed. Their associated properties are presented and the simulation studies have shown significant improvement in the parameter estimations, taking into account the intra-correlation and dependence among the observed mixing random variables. While our proposed model process is guided by a formal fit criteria, Bayesian approach is another option to determine the parameters. However, the proposed approach has the advantage of giving a simple implementation for mixed outcome data.

References

- Balakrishnan, N. & Lai, C. D. (2009). *Continuous bivariate distributions* (2nd ed.). New York: Springer.
- Dennis, J. E. & Schnabel, R. B. (1983). *Numerical methods for unconstrained optimization and nonlinear equations*. Upper Saddle River, NJ: Prentice Hall.
- Diawara, N. (2008). Multivariate gamma distribution based on linear relationship. *Journal of Applied Statistical Science*, 17(4), 1-11.
- Joe, H. (1997). *Multivariate models and dependence concepts*. London: Chapman and Hall.
- Johnson, N., Kotz, S. & Balakrishnan, N. (1997). *Continuous multivariate distributions*. New York: John Wiley.
- Kibble, W. F. (1941). A two-variate gamma type distribution, *Sankhya*, 5, 137-150.
- Kotz, S., Balakrishnan, N. & Johnson, N. L. (2000), *Continuous Multivariate Distributions Volume 1: Models & Applications*. New York: John Wiley.
- Nadarajah, S. (2007). The bivariate gamma exponential distribution with application to drought data. *Journal of Applied Mathematics and Computing*, 24(1-2), 221-230.
- Nadarajah, S. (2009). A bivariate distribution with gamma and beta marginals with application to drought data. *Journal of Applied Statistics*, 36(3), 277-301.
- Taio, G. G. & Cuttman, I. (1965). The inverted Dirichlet distribution with applications. *Journal of the American Statistical Association*, 60(311), 793-805.

Comparison of Estimators in GLM with Binary Data

D. M. Sakate

Shivaji University
Kolhapur, India

D. N. Kashid

Shivaji University
Kolhapur, India

Maximum likelihood estimates (MLE) of regression parameters in the generalized linear models (GLM) are biased and their bias is non negligible when sample size is small. This study focuses on the GLM with binary data with multiple observations on response for each predictor value when sample size is small. The performance of the estimation methods in Cordeiro and McCullagh (1991), Firth (1993) and Pardo et al. (2005) are compared for GLM with binary data using an extensive Monte Carlo simulation study. Performance of these methods for three real data sets is also compared.

Keywords: Binomial regression, modified score function, bias corrected MLE, Minimum ϕ -divergence estimation, Monte Carlo Simulation

Introduction

Generalized linear models (GLM) are frequently used to model small to medium size data. In case of binomial distributed response, logistic regression finds application to model the relationship between response and predictors. Maximum likelihood estimation (MLE) is usually used to fit a logistic regression model. It is well known that under certain regularity conditions, MLE of regression coefficients are consistent and asymptotically normal. However, for finite sample sizes, MLE tend to overestimate with an absolute bias that tends to increase with the magnitude of the parameter and with the ratio of the number of parameters to the number of observations. The bias in MLE decreases with the sample size and goes to zero as sample size tends to infinity. See Byth and McLachlan, (1978), Anderson and Richardson (1979), McLachlan (1980), Pike et al. (1980), Breslow (1981) and Hauck (1984) for the details. As a consequence, methods taking care of bias were explored. Jackknifed MLE and its versions and methods based on approximation

Dr. Sakate is an Assistant Professor in the Department of Statistics. Email him at dms.stats@gmail.com. Dr. Kashid is a Professor in the Department of Statistics. Email him at dnk_stats@unishivaji.ac.in.

of bias using Taylor series expansion are widely studied in the literature. See Bull et al. (1994) and references therein. Cordeiro and McCullagh (1991) proposed second order unbiased MLE in GLM. Further, to simultaneously tackle the problem of bias and separation, Firth (1993) modified the score function to estimate the parameters unbiasedly up to first order. Maiti and Pradhan (2008) empirically established the superiority of these two methods over conditional maximum likelihood estimator in non-separable case through extensive simulation study.

In the last decade, the minimum distance estimators have gained importance in many fields of statistics. Read and Cressie (1988) and Pardo (2006) outlined the use and importance of the ϕ -divergence measures in statistics. Pardo et al. (2005) proposed the minimum ϕ -divergence estimator or minimum distance estimator based on the family of power divergence (Cressie and Read, 1984) characterized by a tuning parameter λ for estimation of regression coefficients in logistic regression. The minimum ϕ -divergence estimator is a generalization of MLE ($\lambda = 0$). Other distance estimators like minimum chi-square estimator ($\lambda = 1$) and minimum Hellinger distance estimator ($\lambda = -1/2$) are particular cases as well. An extensive simulation study in Pardo et al. (2005) and Pardo and Pardo (2008) to choose among the estimators in logistic regression concluded that $2/3$ is a good choice for λ . Hence, minimum ϕ -divergence estimator with $\lambda = 2/3$ emerged as an alternative to MLE in the sense of MSE for small size. The comparison of the minimum distance estimators with those taking care of bias remains the untouched problem of interest.

Estimation in logistic regression

Let Z be a response binary random variable taking value 1 or 0, generally referred to as “success” or “failure” respectively. Let k explanatory variables $\mathbf{x} \in \mathbb{R}^k$ are observed along with the response variable. $\pi(\mathbf{x}) = P(Z = 1 | \mathbf{x} \in \mathbb{R}^k)$ represents the conditional probability, of the value 1 given $\mathbf{x} \in \mathbb{R}^k$. Let X be the $N \times (k + 1)$ matrix with rows $\mathbf{x}_i = (x_{i0}, x_{i1}, \dots, x_{ik})$, $i = 1, \dots, N$ where $x_{i0} = 1$, $\forall i$. The logistic regression model is defined by the conditional probability

$$\pi(\mathbf{x}_i) = \frac{\exp\{\beta_0 + \sum_{j=1}^k \beta_j x_{ij}\}}{1 + \exp\{\beta_0 + \sum_{j=1}^k \beta_j x_{ij}\}}. \quad (1)$$

For more discussion on logistic regression see Hosmer and Lemeshow (1989) and Agresti (1990).

In laboratory or controlled setting, many individuals share same values for their explanatory variables. In other words, for each value of the explanatory variables there are several observed values of the random variable Z . The focus here is on this situation. The notations described earlier are required to be changed slightly. For this, the notations used in Pardo et al. (2005) were followed. Let there be I distinct values of $\mathbf{x}_i = (x_{i0}, x_{i1}, \dots, x_{ik})$, $i = 1, 2, \dots, I$. It is assumed that, for each \mathbf{x}_i , there is a binomial random variable $Y_i \equiv \sum_{j=1}^{n_i} Z_{ij}$ with parameters n_i and $\pi(\mathbf{x}_i)$. The values n_{i1}, \dots, n_{iI} are the observed values of the random variables Y_1, \dots, Y_I , representing the number of successes in n_1, \dots, n_I trials respectively when the explanatory variables are fixed. This divides the entire sample of size N into I subgroups each of size n_i so that $\sum_{i=1}^I n_i = N$. Because Z_i 's are independent, Y_i 's are also independent. Thus, the likelihood function for the logistic regression model is given by

$$L(\beta_0, \dots, \beta_k) = \prod_{i=1}^I \binom{n_i}{n_{i1}} \pi(\mathbf{x}_i^T \boldsymbol{\beta})^{n_{i1}} (1 - \pi(\mathbf{x}_i^T \boldsymbol{\beta}))^{n_i - n_{i1}} \quad (2)$$

The MLE, $\hat{\boldsymbol{\beta}}$ is derived as a solution to score equation

$$l'(\boldsymbol{\beta}) = U(\boldsymbol{\beta}) = \mathbf{0} \quad (3)$$

where $l(\boldsymbol{\beta}) = \log L(\boldsymbol{\beta})$ is the log likelihood function.

Second order bias corrected MLE

As discussed earlier, there are various methods which give rise to bias corrected versions of MLE in logistic regression (Anderson and Richardson, 1979; McLachlan, 1980; Schaefer, 1983; Copas, 1988 and Cordeiro and McCullagh, 1991). Cordeiro and McCullagh (1991) derived an expression for the first order bias using Taylor series expansion. Let $\boldsymbol{\mu}$ be the mean of response variable. The n^{-1} bias of $\hat{\boldsymbol{\beta}}$ in GLM is given by

$$B_1(\boldsymbol{\beta}) = (X^T W X)^{-1} X^T \boldsymbol{\xi}$$

COMPARISON OF ESTIMATORS IN GLM WITH BINARY DATA

where $W = \text{diag}\{V^{-1}\boldsymbol{\mu}'^2\}$ is an $n \times n$ weight matrix and

$$\boldsymbol{\xi} = -(2\psi)^{-1} W^{-1} H_d \mathbf{F} \mathbf{1}.$$

ψ is the dispersion parameter of the GLM, H_d is an $n \times n$ diagonal matrix with the elements of $H = X(X^T W X)^{-1} X^T$ and $F = \text{diag}\{V^{-1}\boldsymbol{\mu}'\boldsymbol{\mu}''\}$ is also an $n \times n$ diagonal matrix. The MLE of $B_1(\boldsymbol{\beta})$ is then subtracted to obtain the second order bias corrected estimate

$$\hat{\boldsymbol{\beta}}^{(2)} = \hat{\boldsymbol{\beta}} - B_1(\hat{\boldsymbol{\beta}}).$$

Modified score function method

The bias in MLE is due to unbiasedness and curvature of score function. The score function is linear in case of normal error regression and hence MLE is unbiased. Firth (1993) modified the score function by introducing small bias in score function. This produces a separation resistant estimator with zero first order bias.

The modified score function is defined as

$$U^*(\boldsymbol{\beta}) = U(\boldsymbol{\beta}) - i(\boldsymbol{\beta}) B_1(\boldsymbol{\beta}) = 0$$

where $i(\boldsymbol{\beta})$ is the Fisher information matrix.

The solution point of the modified score equation locates a stationary point of

$$l^*(\boldsymbol{\beta}) = l(\boldsymbol{\beta}) + \frac{1}{2} \log |i(\boldsymbol{\beta})|$$

or equivalently, of the penalized likelihood function

$$L^*(\boldsymbol{\beta}) = L(\boldsymbol{\beta}) + |i(\boldsymbol{\beta})|^{1/2}$$

where $|i(\boldsymbol{\beta})|^{1/2}$ is the Jeffreys (1946) invariant Prior for the problem.

In GLM with Binary data, the modified score function is (Firth, 1993)

$$U^*(\boldsymbol{\beta}) = U(\boldsymbol{\beta}) - X^T W \boldsymbol{\xi}$$

In logistic regression, $W\xi$ has i^{th} element $h_i (\pi_i - 1/2)$ and h_i is the i^{th} diagonal element of the hat matrix. The solution $U^*(\beta) = 0, \hat{\beta}^{MS}$ is free from the first order bias.

Minimum ϕ -Divergence Estimation

Let $\pi_{i1} = \pi(\mathbf{x}_i^T \beta)$ and $\pi_{i2} = 1 - \pi(\mathbf{x}_i^T \beta)$, $n_{i2} = n_i - n_{i1}$ and $N = \sum_{i=1}^I n_i$. To maximize (2) is equivalent to minimizing the Kullback divergence measure between the probability vectors

$$\hat{\mathbf{p}} = (\hat{p}_{11}, \hat{p}_{12}, \dots, \hat{p}_{I1}, \hat{p}_{I2})^T = \left(\frac{n_{11}}{N}, \frac{n_{12}}{N}, \dots, \frac{n_{I1}}{N}, \frac{n_{I2}}{N} \right)^T \text{ and}$$

$$\mathbf{p}(\beta) = (p_{11}(\beta), p_{12}(\beta), \dots, p_{I1}(\beta), p_{I2}(\beta))^T = \left(\pi_{11} \frac{n_1}{N}, \pi_{12} \frac{n_1}{N}, \dots, \pi_{I1} \frac{n_I}{N}, \pi_{I2} \frac{n_I}{N} \right)^T.$$

MLE for the GLM parameter β can be defined by

$$\hat{\beta} = \arg \min_{\beta \in \Theta} D_{\text{Kullback}}(\hat{\mathbf{p}}, \mathbf{p}(\beta)), \quad (4)$$

where the Kullback divergence measure is given by (see Kullback, 1985)

$$D_{\text{Kullback}}(\hat{\mathbf{p}}, \mathbf{p}(\beta)) = \sum_{j=1}^2 \sum_{i=1}^I \hat{p}_{ij} \log \left(\frac{\hat{p}_{ij}}{p_{ij}(\beta)} \right).$$

This measure is a particular case of the ϕ -divergence defined by Csiszar (1963) and Ali and Silvey (1966),

$$D_{\phi}(\hat{\mathbf{p}}, \mathbf{p}(\beta)) = \sum_{j=1}^2 \sum_{i=1}^I p_{ij}(\beta) \phi \left(\frac{\hat{p}_{ij}}{p_{ij}(\beta)} \right); \phi \in \Phi, \quad (5)$$

where Φ is the class of all convex functions $\phi(x), x > 0$ such that at $x = 1, \phi(1) = \phi'(1) = 0, \phi''(1) > 0$ and at $x = 0, 0 \phi(0/0) = 0$ and $0 \phi(p/0) = p \lim_{u \rightarrow \infty} \phi(u)/u$. For more details, see Vajda (1989) and Pardo (2006).

COMPARISON OF ESTIMATORS IN GLM WITH BINARY DATA

Cressie and Read (1984) introduced an important family of ϕ -divergences called the power divergence family

$$\begin{aligned}\phi_\lambda(x) &= (\lambda(\lambda+1))^{-1} (x^{\lambda+1} - x); \lambda \neq 0, \lambda \neq -1, \\ \phi_0(x) &= \lim_{\lambda \rightarrow 0} \phi_\lambda(x) = x \log(x) - x + 1, \\ \phi_{-1}(x) &= \lim_{\lambda \rightarrow -1} \phi_\lambda(x) = -\log(x) - x - 1.\end{aligned}\tag{6}$$

It is interesting to note that

$$D_{\phi_0}(\hat{p}, p(\beta)) = D_{Kullback}(\hat{p}, p(\beta)).\tag{7}$$

The minimum ϕ -divergence estimator (Pardo et al., 2005) in logistic regression is given by

$$\hat{\beta}_\phi = \arg \min_{\beta \in \Theta} D_\phi(\hat{p}, p(\beta))\tag{8}$$

Methodology

Performance comparison for real data

Usefulness of any method can be established only when it can be applied to a real data. To this end, the performance of these methods is compared for prediction purposes when real data is used. Three real data sets are employed as examples. The numbers of predictors in each real data are not same. A single predictor, two predictor and multiple predictor situations are considered in real data to compare the prediction performance of the methods.

Example 1: Single Predictor Case

First, for the single predictor case, pneumoconiosis data (Montgomery et. al., 2006; pp. 449) concerning the proportion of coal miners who exhibit symptoms of severe pneumoconiosis and the numbers of years of exposure (X_1) is analyzed. The data includes $n = 8$ observations on number of years of exposure, number of severe cases and total number of miners. Table 1 presents the estimated regression coefficients using the methods discussed earlier.

Table 1. Estimates of regression coefficients: Example 1

Predictors	$\hat{\beta}$	$\hat{\beta}^{(2)}$	$\hat{\beta}_{\phi}$	$\hat{\beta}^{MS}$
Intercept	-4.55621	-4.47047	-4.54838	-4.48677
# Years of Exposure	0.07889	0.07707	0.08753	0.07747

Example 2: Two Predictors Case

The two predictors case example uses ingots data, which appeared in Cox and Snell (1989, pp. 10-11). This data is a result of an industrial experiment concerning steel ingots. It consists of 19 observations on the number of ingots not ready for rolling out of certain number of trials tested for a number of heating time (X_1) and soaking time (X_2). This data is also analyzed by Pardo et al. (2005) to illustrate the use of minimum ϕ -divergence estimator. Table 2 presents the estimates of regression coefficients in the binomial logistic regression.

Table 2. Estimates of regression coefficients: Example 2

Predictors	$\hat{\beta}$	$\hat{\beta}^{(2)}$	$\hat{\beta}_{\phi}$	$\hat{\beta}^{MS}$
Intercept	-5.51316	-5.42360	-4.88651	-5.47685
Heating time	0.07688	0.07573	0.06881	0.07690
Soaking time	0.07201	0.11485	0.04469	0.10876

Example 3: Multiple Predictors Case

In this next example, a real data with more than two predictors is analyzed and considers data (Andersen 1997, pp. 171) used by Pardo and Pardo (2008) to illustrate the variable selection method based on minimum ϕ -divergence estimator. The data consists of observations on 6 objective indicators (X_1, \dots, X_6) of the actual indoor climate in 10 classrooms of a Danish Institute, the number of students in the class and the number of yes-answers to the question whether they felt that the indoor climate at the moment was pleasant or not so pleasant. Table 3 presents the estimates of regression coefficients in the binomial logistic regression. The minimum ϕ -divergence estimate of coefficient of X_3 and X_5 differ in magnitude as compared to estimates from other three methods to a larger extent.

COMPARISON OF ESTIMATORS IN GLM WITH BINARY DATA

Table 3. Estimates of regression coefficients: Example 3

Predictors	$\hat{\beta}$	$\hat{\beta}^{(2)}$	$\hat{\beta}_{\phi}$	$\hat{\beta}^{MS}$
Intercept	5.75029	5.48737	7.15570	5.50296
X_1	0.53530	0.51090	1.29870	0.51380
X_2	-0.51320	-0.49247	-1.15230	-0.49442
X_3	9.04758	8.65945	19.28910	8.69477
X_4	0.64191	0.61565	1.42370	0.61780
X_5	8.93732	8.67950	25.20410	8.67973
X_6	-0.04478	-0.04263	-0.07700	-0.04283

It is not possible to decide between the estimators merely by looking at the estimated regression coefficients. To compare the performance in each example, the predicted probabilities of success using each estimator were computed. As a measure of discrepancy between estimated and observed probability of success, the Pearson chi-square defined as

$$\text{Pearson chi-square} = \sum_{i=1}^I \left(\frac{P_i - \hat{\pi}_{i1}^c}{\sqrt{n_i \hat{\pi}_{i1}^c \hat{\pi}_{i2}^c}} \right)^2$$

was used, where, $P_i = \frac{n_{i1}}{n_i}$ and c represents the method used to estimate probability of success. The Pearson chi-square corresponding to estimators considered for all the examples are listed in the Table 4. The Pearson chi-square for $\hat{\beta}_{\phi}$ is smallest in case of Examples 1 and 3. For Example 2, value of Pearson chi-square is smallest for $\hat{\beta}$. It reveals that the performance of minimum ϕ -divergence estimators to predict probabilities of success in binomial logistic regression applied to real data is better than MLE and its bias corrected versions.

Table 4. Pearson chi-square

Estimator	$\hat{\beta}$	$\hat{\beta}^{(2)}$	$\hat{\beta}_{\phi}$	$\hat{\beta}^{MS}$
Example 1	0.0058	0.0057	0.0056	0.0019
Example 2	0.7570	0.8733	0.8648	0.8682
Example 3	0.2419	0.2393	0.2394	0.0216

Monte-Carlo Simulation Study

Estimation methods were compared using Monte-Carlo simulation; a two predictor binomial logistic regression model including an intercept was considered. The design matrix is of order 11×3 with first column as ones. The other two columns contain random numbers from two independent standard uniform distributions. To generate observations on response variable, two different parameter structures were considered and accordingly two different models were defined as Model I: $\beta = (1, 2, -3)$; Model II: $\beta = (1, -1.5, 2)$.

Table 5. AMSE with its SD

		$\hat{\beta}$	$\hat{\beta}^{(2)}$	$\hat{\beta}_{\phi}$	$\hat{\beta}^{MS}$
Model I	n^1	2.4014 (0.5287)	2.4195 (0.5235)	2.4173 (0.5244)	0.2902 (0.3909)
	n^2	2.4055 (0.6030)	2.4355 (0.5951)	2.4298 (0.5956)	0.4202 (0.5254)
	n^3	2.3944 (0.4684)	2.4068 (0.4646)	2.4055 (0.4654)	0.1768 (0.2292)
	n^4	2.4228 (0.5837)	2.4553 (0.5739)	2.4512 (0.5759)	0.4210 (0.4849)
	n^5	2.4455 (0.7145)	2.4945 (0.6952)	2.4873 (0.6992)	0.7497 (1.0160)
	n^6	2.4371 (0.7172)	2.4811 (0.6996)	2.4757 (0.7032)	0.7200 (0.9379)
	n^7	2.4201 (0.5828)	2.4518 (0.5755)	2.4464 (0.5761)	0.4432 (0.5208)
	n^8	2.4332 (0.5633)	2.4562 (0.5566)	2.4537 (0.5580)	0.3410 (0.3640)
	n^9	2.5535 (0.9460)	2.6306 (0.8986)	2.6189 (0.9091)	1.5971 (5.8009)
	n^{10}	2.6685 (1.1593)	2.7568 (1.0321)	2.7364 (1.0605)	3.1885 (6.3671)

COMPARISON OF ESTIMATORS IN GLM WITH BINARY DATA

Table 5, continued

		$\hat{\beta}$	$\hat{\beta}^{(2)}$	$\hat{\beta}_{\phi}$	$\hat{\beta}^{MS}$
Model II	n^1	1.8489 (0.2223)	1.8528 (0.2206)	1.8524 (0.2207)	0.3023 (0.3825)
	n^2	1.8673 (0.2760)	1.8732 (0.2722)	1.8723 (0.2725)	0.5364 (0.7490)
	n^3	1.8511 (0.1868)	1.8537 (0.1857)	1.8535 (0.1858)	0.2053 (0.3301)
	n^4	1.8588 (0.2877)	1.8653 (0.2828)	1.8648 (0.2833)	0.5813 (0.8713)
	n^5	1.8819 (0.3377)	1.8913 (0.3297)	1.8903 (0.3305)	0.8732 (1.1947)
	n^6	1.8753 (0.3233)	1.8840 (0.3165)	1.8833 (0.3172)	0.7925 (0.9626)
	n^7	1.8644 (0.2984)	1.8705 (0.2935)	1.8695 (0.2942)	0.5432 (0.6855)
	n^8	1.8500 (0.2526)	1.8550 (0.2498)	1.8547 (0.2501)	0.4296 (0.8013)
	n^9	1.9098 (0.4462)	1.9237 (0.4260)	1.9221 (0.4284)	1.8013 (3.5744)
	n^{10}	1.9155 (0.4511)	1.9282 (0.4276)	1.9268 (0.4305)	2.1006 (7.0825)

Table 6. Average absolute bias

		$\hat{\beta}$	$\hat{\beta}^{(2)}$	$\hat{\beta}_{\phi}$	$\hat{\beta}^{MS}$
Model I	n^1	0.2556	0.2534	0.2535	0.0023
	n^2	0.2580	0.2539	0.2545	0.0048
	n^3	0.2546	0.2531	0.2531	0.0018
	n^4	0.2550	0.2508	0.2511	0.0002
	n^5	0.2576	0.2508	0.2513	0.0065
	n^6	0.2640	0.2575	0.2580	0.0108
	n^7	0.2586	0.2543	0.2548	0.0055
	n^8	0.2516	0.2487	0.2488	0.0094
	n^9	0.2606	0.2471	0.2485	0.0210
	n^{10}	0.2721	0.2475	0.2510	0.0010
Model II	n^1	0.5785	0.5780	0.5780	0.0075
	n^2	0.5793	0.5785	0.5786	0.0149
	n^3	0.5780	0.5777	0.5777	0.0031
	n^4	0.5788	0.5779	0.5779	0.0159
	n^5	0.5792	0.5778	0.5779	0.0258
	n^6	0.5787	0.5774	0.5774	0.0120
	n^7	0.5782	0.5773	0.5774	0.0093
	n^8	0.5783	0.5776	0.5776	0.0079
	n^9	0.5842	0.5813	0.5814	0.0309
	n^{10}	0.5810	0.5781	0.5782	0.0182

Consider the following 10 different combinations of number of trials

$$\begin{aligned}
 n^1 &: 15, 15, 15, 15, 30, 30, 30, 30, 40, 40, 40 \\
 n^2 &: 5, 5, 5, 5, 15, 15, 15, 15, 40, 40, 40 \\
 n^3 &: 40, 40, 40, 40, 40, 40, 40, 40, 40, 40, 40 \\
 n^4 &: 10, 10, 10, 10, 20, 20, 20, 20, 15, 15, 15 \\
 n^5 &: 10, 10, 10, 10, 5, 5, 5, 5, 15, 15, 15 \\
 n^6 &: 10, 10, 10, 10, 10, 10, 10, 10, 10, 10, 10 \\
 n^7 &: 5, 5, 5, 5, 30, 30, 30, 30, 15, 15, 15, 15 \\
 n^8 &: 20, 20, 20, 20, 20, 20, 20, 20, 20, 20, 20, 20 \\
 n^9 &: 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5, 5 \\
 n^{10} &: 5, 5, 5, 5, 5, 5, 5, 5, 4, 4, 4
 \end{aligned}$$

The observations on response variable are random numbers from $B(n_i, \pi_{i1})$. In this way, 20 models were generated differing in parameter structure and structure of number of trials. Unknown regression coefficients were estimated using four methods including MLE. To compute the minimum ϕ -divergence estimate, the power divergence family in (6) with $\lambda = 2/3$ were used as suggested in Pardo et. al. (2005). Each model was simulated 1,000 times and average MSE (AMSE) and average absolute bias in estimate due to each estimation method are reported in the Tables 5 and 6. The figures in parentheses represent standard deviation (SD) of MSE. The AMSE and average absolute bias were computed using the following formulae

$$\begin{aligned}
 \text{AMSE} &= \frac{1}{1000} \sum_{i=1}^{1000} \frac{1}{3} \sum_{j=0}^2 (\hat{\beta}_{ij}^c - \beta_j)^2. \\
 \text{Average absolute bias} &= \frac{1}{1000} \sum_{i=1}^{1000} \frac{1}{3} \sum_{j=0}^2 |\hat{\beta}_{ij}^c - \beta_j|.
 \end{aligned}$$

Results

It is evident from the Tables 5 and 6, the minimum ϕ -divergence estimator has smaller MSE and bias as compared to others for all combinations of number of trials. For a small magnitude of number of trials, as in case of last combination, the

COMPARISON OF ESTIMATORS IN GLM WITH BINARY DATA

AMSE of all the estimators is more or less same; however, variability in the minimum ϕ -divergence estimate is quite high. The estimate based on modified score function and second order bias corrected MLE are close enough to uncorrected MLE in this setting. The bias correction obtained as such is negligible. From Tables 5 and 6, performance of minimum ϕ -divergence estimator is better than the others for all but last two combinations of number of trials i.e., n^9 and n^{10} .

The variation in MSE and absolute bias averaged over the three regression coefficients is shown in Figures 1 and 2 respectively for n^1 and n^{10} . Although the motivation behind defining the minimum ϕ -divergence estimator was altogether different, it performs better than the bias corrected versions of MLE. This makes the minimum ϕ -divergence estimator an attractive alternative to MLE as well as its bias corrected versions in binomial logistic regression.

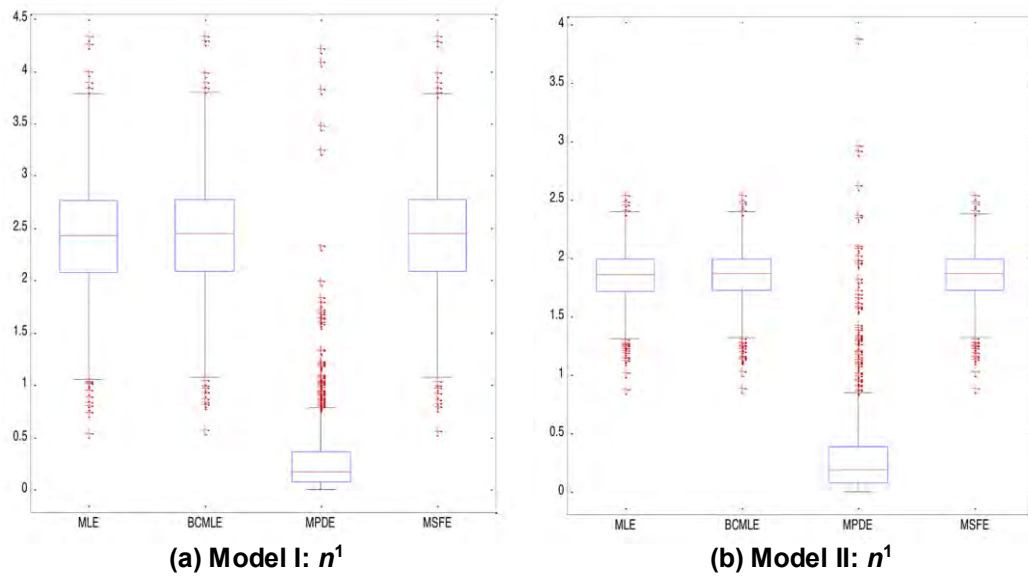


Figure 1 (a, b). Box Plot of MSE of estimates averaged over three regression coefficients.

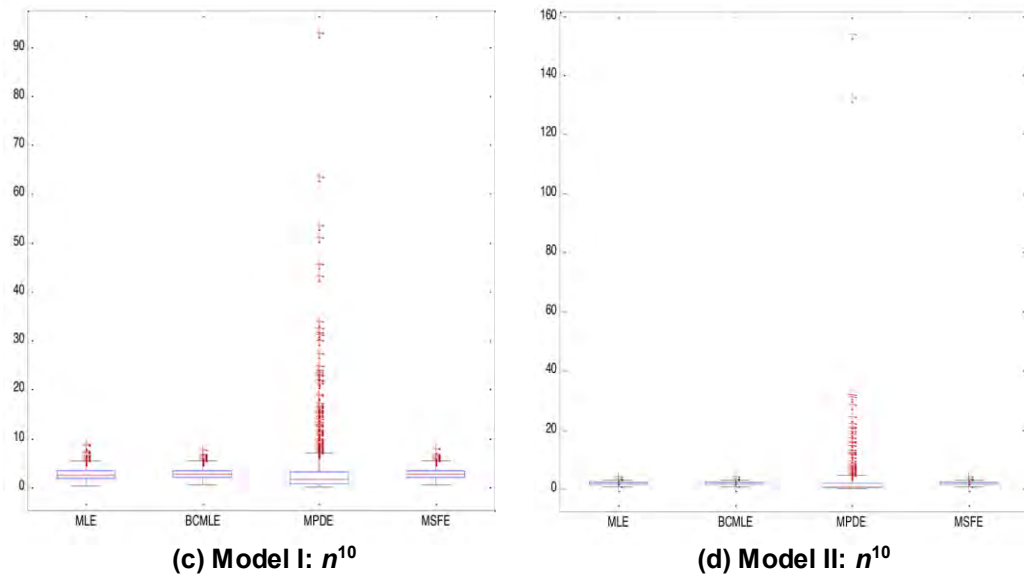


Figure 1 (c, d). Box Plot of MSE of estimates averaged over three regression coefficients.

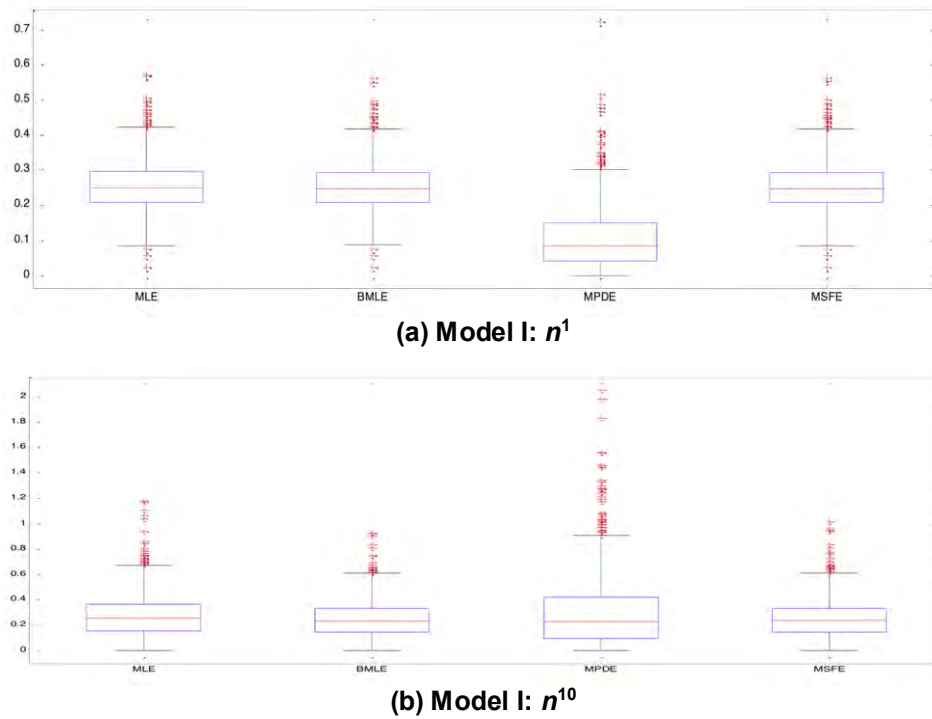


Figure 2 (a, b). Box Plot of bias of estimates averaged over three regression coefficients.

COMPARISON OF ESTIMATORS IN GLM WITH BINARY DATA

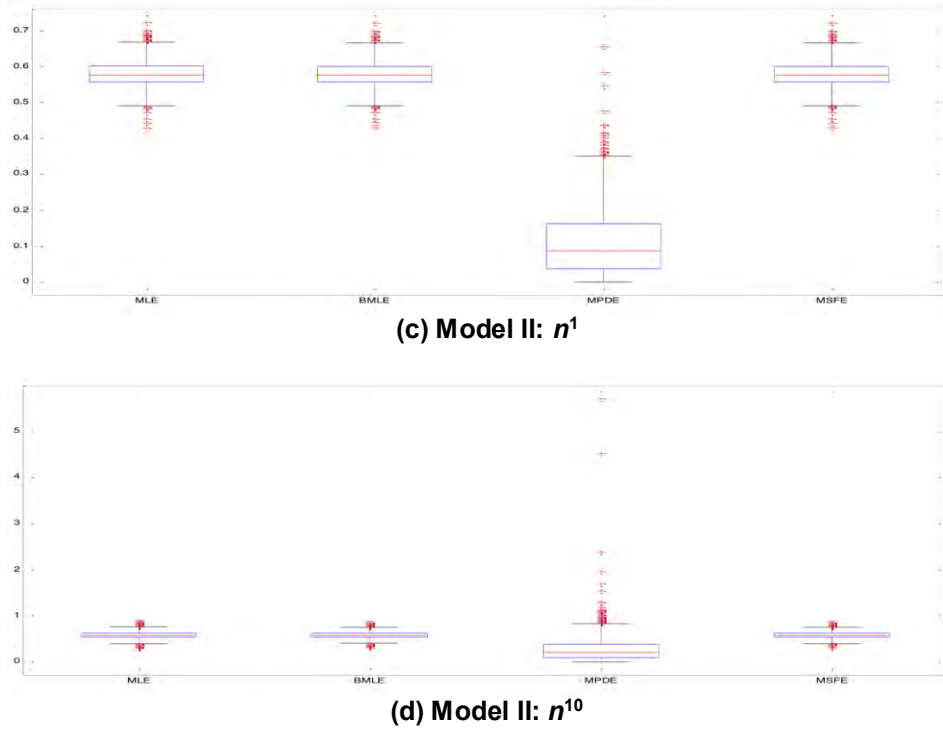


Figure 2 (c, d). Box Plot of bias of estimates averaged over three regression coefficients.

Conclusion

The performance of some of the estimators belonging to two different classes, i.e., minimum distance estimators and bias corrected MLE in a binomial logistic regression model, was compared. Three real data examples from different fields followed by a Monte Carlo simulation study were used to illustrate the comparisons. Results show that second order bias corrected MLE and estimates obtained using modified score function method lead to an estimate, which is same as MLE when number of trials is large. From this comparison study it may be concluded that, for a number of trials greater than 5, minimum ϕ -divergence estimator is an attractive alternative to MLE as well as bias corrected and modified score function method.

References

- Agresti, A. (1990). *Categorical data analysis*. New York: John Wiley and Sons.
- Ali, S. M. & Silvey, S. D. (1966). A general class of coefficients of divergence of one distribution from another. *Journal of the Royal Statistical Society, Series B*, 26, 131-142.
- Andersen, E. B. (1997). *Introduction to the statistical analysis of categorical data I*. New York: Springer.
- Anderson, J. A. & Richardson, S. C. (1979). Logistic discrimination and bias correction in maximum likelihood estimation. *Technometrics*, 21, 71-78.
- Breslow, N. (1981). Odds ratio estimators when the data are sparse. *Biometrika*, 68, 73-84.
- Bull, S. S., Hauck, W. W. & Greenwood, C. M. T. (1994). Two-step Jackknife bias reduction for logistic regression MLEs. *Communication in Statistics – Simulation and Computation*, 23, 59-88.
- Byth, K. & McLachlan, G. J. (1978). The biases associated with maximum likelihood methods of estimation of the multivariate logistic risk function. *Communication in Statistics – Theory and Methods*, A7, 877-890.
- Copas, J. B. (1988). Binary regression models for contaminated data with discussion. *Journal of the Royal Statistical Society, Series B*, 50, 225-265.
- Cordeiro, G. M. & McCullagh, P. (1991). Bias correction in generalized linear models. *Journal of the Royal Statistical Society, Series B*, 53, 629-643.
- Cox, D. R. & Snell, E. J. (1989). *Analysis of Binary Data*. London: Chapman and Hall.
- Cressie, N. A. C. & Read, T. R. C. (1984). Multinomial goodness of fit tests. *Journal of the Royal Statistical Society, Series B*, 46, 440-464.
- Csiszár, I. (1963). Eine Informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität Markoffscher Ketten. *Publications of the mathematical Institute of Hungarian Academy of Sciences, Series A*, 8, 84–108.
- Firth, D. (1993) Bias Reduction of Maximum Likelihood Estimates. *Biometrika*, 80, 27-38.
- Hauck, W. W. (1984). A comparative study of conditional maximum likelihood estimation of a common odds ratio. *Biometrics*, 40, 1117-1123.

COMPARISON OF ESTIMATORS IN GLM WITH BINARY DATA

Hosmer, D. W. & Lemeshow, S. (1989). *Applied Logistic Regression*. New York: John Wiley and Sons.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London* 186, 453–461.

Kullback, S. (1985). ‘Kullback information.’ In S. Kotz & N. L. Johnson (Eds.). *Encyclopedia of Statistical Sciences*, 4, 421–425. New York: John Wiley and Sons.

Maiti, T. & Pradhan, V. (2008). A comparative study of the bias corrected estimates in logistic regression. *Statistical Methods in Medical Research*, 17, 621–634.

McCullagh P. & Nelder J. A. (1989). *Generalized Linear models* (Second ed.). London: Chapman and Hall.

McLachlan, G. J. (1980). A note on bias correction in maximum likelihood estimation with logistic discrimination. *Technometrics*, 22, 621–627.

Montgomery, D. C., Peck, E. A & Vining, G. G. (2006). *Introduction to linear regression analysis*. New York: John Wiley and Sons.

Pardo, J. A., Pardo, L., & Pardo, M. C. (2005). Minimum ϕ -divergence estimator in logistic regression models. *Statistical Papers*, 47, 91–108.

Pardo, J. A. & Pardo, M. C. (2008). Minimum ϕ -divergence estimator and ϕ -divergence statistics in generalized linear models with binary data. *Methodology and Computing in Applied Probability*, 10, 357–379.

Pardo, L. (2006). *Statistical inference based on divergence measures*. New York: Taylor and Francis Group, LLC.

Pike, M. C., Hill, A. P., & Smith, P. G. (1980). Bias and efficiency in logistic analysis of stratified case control studies. *International Journal of Epidemiology*, 9, 705–724.

Read, T. R. C. & Cresie, N. (1988). *Goodness of fit statistics for discrete multivariate data*. New York: Springer.

Schaefer, R. L. (1983). Bias correction in maximum likelihood logistic regression, *Statistics in Medicine*, 2, 71–78.

Vajda, I. (1989). *Theory of statistical inference and information*. Boston: Kluwer.

Gumbel-Weibull Distribution: Properties and Applications

Raid Al-Aqtash
Marshall University
Huntington, WV

Carl Lee
Central Michigan University
Mount Pleasant, MI

Felix Famoye
Central Michigan University
Mount Pleasant, MI

Some properties of the Gumbel-Weibull distribution including the mean deviations and modes are studied. A detailed discussion of regions of unimodality and bimodality is given. The method of maximum likelihood is proposed for estimating the distribution parameters and a simulation is conducted to study the performance of the method. Three tests are given for testing the significance of a distribution parameter. The applications of Gumbel-Weibull distribution are emphasized. Five data sets are used to illustrate the flexibility of the distribution in fitting unimodal and bimodal data sets.

Keywords: Mean deviation, bimodality, maximum likelihood estimation, lifetime data

Introduction

Problems on extreme values appeared in the work of Nicholas Bernoulli back in 1709 for studying the problem of the mean largest distance from origin for n random numbers on a straight line (see Johnson et al., 1995, p. 1). During 1920s and 1930s, many papers on the distribution of extremes appeared in the literature. Gumbel (1958) gave detailed results on extreme value theory in his book *Statistics of Extremes*. Furthermore, Gumbel (1958) has been referred to by Johnson et al. (1995) as the first to bring attention to the possibility of using the Gumbel distribution to model extreme values of random data. For more information on extreme value distributions, see Johnson et al. (1995), Gumbel (1958), Kotz and Nadarajah (2000), and Beirlant et al. (2006).

The Weibull distribution is well known for its ability to model different types of data. Weibull distribution also has many applications in risk analysis and quality

Dr. Al-Aqtash is an Assistant Professor in the Department of Mathematics. Email him at alaqtash@marshall.edu. Dr. Lee is a Professor in the Department of Mathematics. Email him at: carl.lee@cmich.edu. Dr. Famoye is a Professor in Department of Mathematics. Email him at: felix.famoye@cmich.edu.

GUMBEL-WEIBULL DISTRIBUTION

control because its hazard rate is decreasing when the shape parameter $a < 1$, constant when $a = 1$, and increasing when $a > 1$. For more information on Weibull distribution, see Johnson et al. (1994).

Alzaatreh, Lee and Famoye (2013) proposed a method for generating new distributions, namely, the T - X family. The cumulative distribution function (CDF) of the T - X family is defined as $G(x) = \int_{-\infty}^{W(F(x))} r(t) dt = R(W(F(x)))$, where X is any continuous random variable with CDF $F(x)$ and probability density function (PDF) $f(x)$, $r(t)$ and $R(t)$ are the PDF and the CDF of a continuous random variable T . They further studied a T - X family by defining $W(F(x)) = -\log(1-F(x))$, where the random variable T is defined on $(0, \infty)$. Al-Aqtash et al. (2014) defined and studied a family of T - X distributions arising from the logit function $W(F(x)) = \ln\{F(x)/(1-F(x))\}$ and provided some general properties of this T - X family, including symmetry, quantile function and Shannon entropy. The CDF of the T - X distribution has the form $G(x) = \int_{-\infty}^{\ln\{F(x)/(1-F(x))\}} r(t) dt = R(\ln\{F(x)/(1-F(x))\})$, where T is defined on $(-\infty, \infty)$. The main difference from the T - X family studied by Alzaatreh, Lee and Famoye (2013) is that the random variable T is defined on $(-\infty, \infty)$ instead of $(0, \infty)$. Taking T to be Gumbel with PDF $r(t) = (1/\sigma) e^{-(t-\nu)/\sigma} \exp\{-e^{-(t-\nu)/\sigma}\}$ and X to be Weibull with PDF $f(x) = (a/\lambda) (x/\lambda)^{a-1} \exp\{-(x/\lambda)^a\}$, Al-Aqtash et al. (2014) defined the four-parameter Gumbel-Weibull distribution (GWD) with CDF

$$G(x) = \exp\left\{-e^{\nu/\sigma} \left(e^{(x/\lambda)^a} - 1\right)^{-1/\sigma}\right\} = \exp\left\{-\beta \left(e^{(x/\lambda)^a} - 1\right)^{-1/\sigma}\right\},$$

where $0 < x < \infty$, $\beta = e^{\nu/\sigma}$, $\sigma, a, \lambda > 0$ and $-\infty < \nu < \infty$. The corresponding PDF is

$$g(x) = \frac{a\beta}{\lambda\sigma} (x/\lambda)^{a-1} e^{(x/\lambda)^a} \left(e^{(x/\lambda)^a} - 1\right)^{-1-1/\sigma} \exp\left\{-\beta \left(e^{(x/\lambda)^a} - 1\right)^{-1/\sigma}\right\}. \quad (1)$$

Al-Aqtash et al. (2014) derived some properties of GWD such as moments and Shannon entropy. In this article, additional properties including the mean deviations and modality are studied. In particular, the applications of this distribution are emphasized.

Mean deviations

Two measures of spread, namely the mean deviation from the mean and the mean deviation from the median for GWD are now presented. Let X be a random variable from the GWD with mean μ and median M .

The mean deviation from the mean is given by

$$\begin{aligned} E(|X - \mu|) &= \int_0^{\infty} |x - \mu| g(x) dx = 2 \int_0^{\mu} (\mu - x) g(x) dx + \int_0^{\infty} (x - \mu) g(x) dx \\ &= 2\mu G(\mu) - 2 \int_0^{\mu} xg(x) dx, \end{aligned} \quad (2)$$

where $G(\mu) = \int_0^{\mu} g(x) dx$. In a similar way, the mean deviation from the median is given by

$$\begin{aligned} E(|X - M|) &= \int_0^{\infty} |x - M| g(x) dx \\ &= 2 \int_0^M (M - x) g(x) dx + \int_0^{\infty} (x - M) g(x) dx \\ &= \mu - 2 \int_0^M xg(x) dx. \end{aligned} \quad (3)$$

The integral $\int_0^v xg(x) dx$ in (2) and (3) can be computed numerically.

Modes of Gumbel-Weibull distribution

Al-Aqtash et al. (2014) studied the moments and skewness in detail, and mentioned that GWD can be unimodal or bimodal. However, no study on GWD modes was given. The modes of GWD are provided and the regions of unimodality and bimodality for specific values of β are investigated. Differentiating the density of GWD in (1) with respect to x results in

$$\begin{aligned} g'(x) &= \frac{a\beta}{\lambda^2\sigma} \left((x/\lambda)^{a-2} e^{(x/\lambda)^a} \left(e^{(x/\lambda)^a} - 1 \right)^{-2-1/\sigma} \exp \left\{ -\beta \left(e^{(x/\lambda)^a} - 1 \right)^{-1/\sigma} \right\} \right) \\ &\quad \times \left[(a-1) \left(e^{(x/\lambda)^a} - 1 \right) - a(x/\lambda)^a + (a/\sigma)(x/\lambda)^a e^{(x/\lambda)^a} \left\{ \beta \left(e^{(x/\lambda)^a} - 1 \right)^{-1/\sigma} - 1 \right\} \right]. \end{aligned}$$

GUMBEL-WEIBULL DISTRIBUTION

The derivative $g'(x)$ does not exist when $x = 0$. Other critical point(s) satisfy $g'(x) = 0$, hence if there is a mode for GWD, then it will be either at $x = 0$ or it will satisfy equation

$$(a-1)\left(e^{(x/\lambda)^a} - 1\right) - a(x/\lambda)^a + (a/\sigma)(x/\lambda)^a e^{(x/\lambda)^a} \left\{ \beta \left(e^{(x/\lambda)^a} - 1 \right)^{-1/\sigma} - 1 \right\} = 0. \quad (4)$$

In (4), set $\lambda = 1$ because λ is a scale parameter and it will not affect the distribution shape. The expression on the left hand side of (4) simplifies to

$$(a-1)\left(e^{x^a} - 1\right) - ax^a + (a/\sigma)x^a e^{x^a} \left\{ \beta \left(e^{x^a} - 1 \right)^{-1/\sigma} - 1 \right\}. \quad (5)$$

Analytical solution of equating (5) to zero is not possible. Numerical approximation is applied to study the modes and the regions of unimodality and bimodality. To study the modes of the GWD, fix β and allow a and σ to change from 0.01 to 10 at an increment of 0.01. This gives one million different ordered pairs (σ, a) . For each ordered pair, (5) is computed as a vector using a vector x with values 0.000001, 0.00001, 0.0001, 0.001, and then from 0.01 to 10 at an increment of 0.01. Values of x below 0.000001 cause execution errors as a approaches 10, because $\left(e^{x^a} - 1\right)^{-1/\sigma}$ grows rapidly toward ∞ as x approaches zero. Because it was observed that the PDF of GWD quickly approaches zero when $x > 10$, x less than 10 was chosen.

A 1000×1000 matrix P was constructed to store the symbols $\{-2, -1, 1, 2\}$, and track how many times the sign of the components of the vector (5) changes. If the vector is always negative, set $P_{(\sigma,a)} = -1$ indicating one mode at point $x = 0$. If (5) starts negative, then changes to positive and finally becomes negative, set $P_{(\sigma,a)} = -2$ indicating one mode at point $x = 0$ and another mode at $x > 0$. If (5) starts positive and becomes negative, then changes to positive, and finally becomes negative, set $P_{(\sigma,a)} = 2$ indicating two distinct modes at $x > 0$. Finally, if (5) starts positive, then becomes negative, set $P_{(\sigma,a)} = 1$ indicating one mode at $x > 0$.

For fixed β , the parameter space of GWD is made up of four regions, each region contains a symbol from $\{-1, -2, 2, 1\}$, separated by boundary curves. The boundary curves are estimated by using regression model. The four distinct regions for GWD modes are marked R1, R2, R3, and R4 corresponding to symbols $-1, -2, 2$ and 1 respectively. The regions R1 and R4 determine the region of unimodality, while the regions R2 and R3 determine the region of bimodality. Figures 1, 2 and

3 show the regions of unimodality and bimodality are not robust to the parameter β . To save space the following three cases are presented to demonstrate the complexity of the modal property of GWD.

Case 1 ($\beta = 0.5$)

Figure 1 shows the four distinct regions R1, R2, R3, and R4 for GWD modes, and four PDFs representing the four regions when $\lambda = 1$. For example, the PDF of GWD with parameters $\sigma = 10$ and $a = 1.7$ falls in R2, thus it is bimodal with one mode at 0 and the other mode is at $x = 2.71$. In this case, only two curves C1 and C2 are found as boundaries for the four regions. The regression models for C1 and C2 have $R^2 = 100\%$ and are given by

$$C1: \sigma = -0.04373 + 5.556a - 0.1453a^2 + 0.04509a^3, a < 1.85, \text{ and}$$

$$C2: \sigma = (0.6539 - 0.5853a^{-1} + 0.2609a^{-2} - 0.2292a^{-3})^{-2}, 1.52 < a < 10.$$

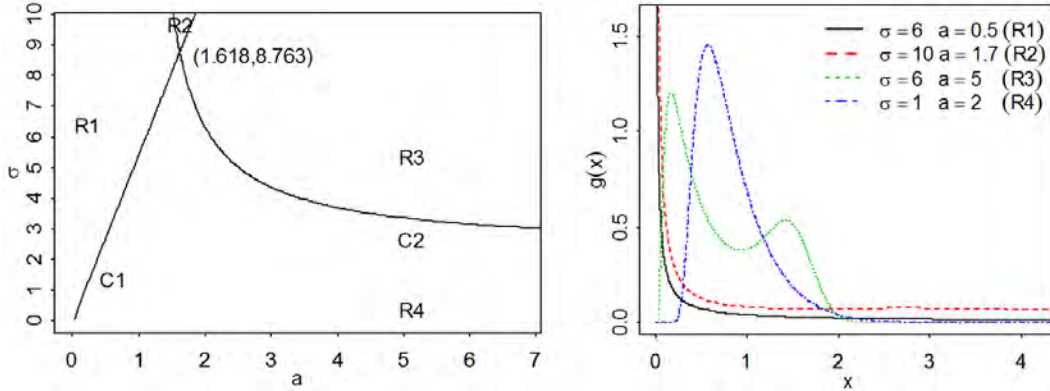


Figure 1. Regions of unimodality and bimodality and PDFs of GWD when $\beta=0.5$ and $\lambda=1$

Case 2 ($\beta = 1$)

Figure 2 shows the four distinct regions R1, R2, R3, and R4 for GWD modes, and four PDFs from the four regions when $\lambda = 1$. For example, the PDF corresponding to $\sigma = 10$ and $a = 1.4$ is bimodal, the first mode is at 0 and the other mode is at $x = 3.55$. Three curves C1, C2, and C3 separate the four regions. The regression models for C1, C2, and C3 all have $R^2 = 100\%$ and are given by

GUMBEL-WEIBULL DISTRIBUTION

$$\begin{aligned}
 C1: & \sigma = -0.05499 + 6.977a - 0.2862a^2 + 0.1072a^3, a < 1.49, \\
 C2: & \sigma = \left(-1.668 + 2.999a^{1/2} - 1.655a + 0.3443a^{3/2}\right)^{-1}, 1.27 < a < 3.26, \text{ and} \\
 C3: & \sigma = -12.26 + 14.43a^{1/2} - 4.332a + 0.5391a^{3/2}, 3.26 < a < 10.
 \end{aligned}$$

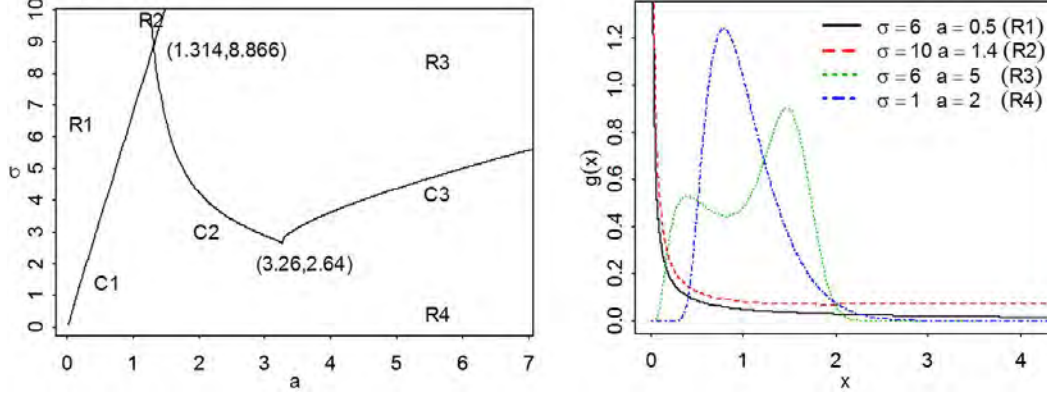


Figure 2. Regions of unimodality and bimodality and PDFs of GWD when $\beta=1$ and $\lambda=1$

Case 3 ($\beta = 2$)

Figure 3 shows the four distinct regions R1, R2, R3, and R4 for GWD modes separated by three curves and four PDFs from the four regions when $\lambda = 1$. For example, the PDF corresponding to $\sigma = 10$ and $a = 1$ is bimodal with modes at 0 and at $x = 6.83$. The three regression models, all with $R^2 = 100\%$, are given by

$$\begin{aligned}
 C1: & \sigma = \left(-0.01444 + 25.78a^{3/2} + 0.1418a^3 - 0.1129a^{9/2}\right)^{2/3}, a < 1.15, \\
 C2: & \sigma = \left(-0.9390 + 2.688a^2 - 1.814a^4 + 0.4480a^6\right)^{-2}, 0.93 < a < 1.289, \text{ and} \\
 C3: & \sigma = 0.3479 + 1.531a - 0.1226a^2 + 0.004977a^3, 1.289 < a < 5.05.
 \end{aligned}$$

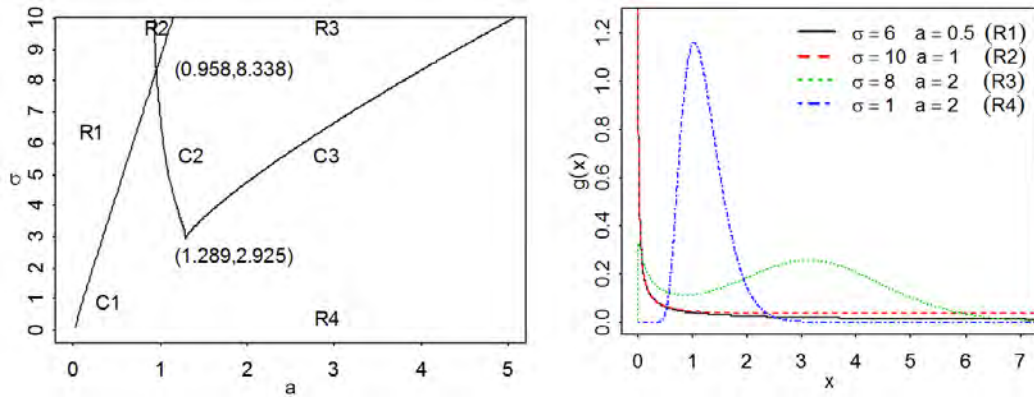


Figure 3. Regions of unimodality and bimodality and PDFs of GWD when $\beta=2$ and $\lambda=1$

The distance between the two modes when GWD is bimodal was also examined, which is demonstrated in Table 1. From Table 1 it is observed that, when σ increases and all the other parameters are fixed, the distance between two modes increases, and when a increases, the distance decreases. However, no clear increasing or decreasing pattern is observed as β increases when all other parameters are fixed. Figure 4 further illustrates the diverse shapes of GWD and the distances between two modes when σ , β , or a changes, respectively.

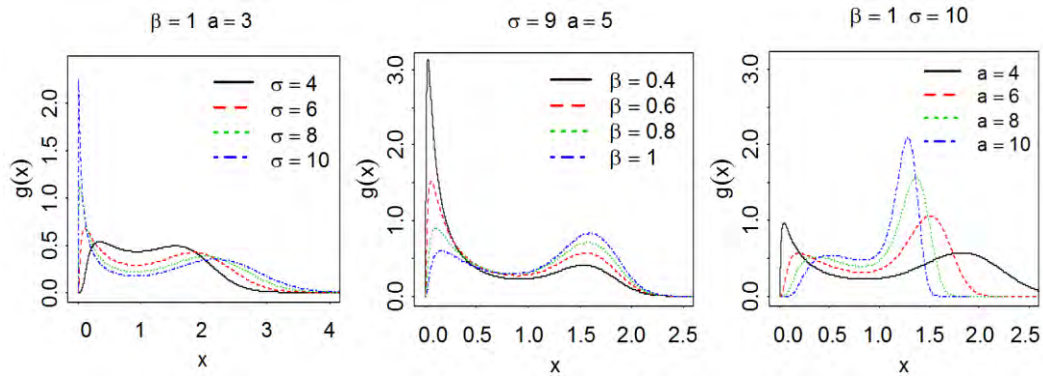


Figure 4. PDFs of GWD when the distribution is bimodal and $\lambda=1$

GUMBEL-WEIBULL DISTRIBUTION

Table 1. Distance between two modes when GWD is bimodal and $\lambda = 1$

a	σ	$\beta = 0.6$	$\beta = 0.8$	$\beta = 1$	$\beta = 1.5$	$\beta = 2$
2	4				1.68	
	5		1.62	1.91	2.13	2.14
	6	1.82	2.07	2.21	2.44	2.57
	7	2.14	2.29	2.43	2.68	2.87
	8	2.34	2.49	2.61	2.89	3.10
	9	2.51	2.65	2.79	3.07	3.30
	10	2.65	2.80	2.94	3.25	3.49
3	4		1.23	1.20		
	5	1.50	1.52	1.51	1.35	
	6	1.68	1.70	1.71	1.67	
	7	1.81	1.85	1.87	1.88	1.78
	8	1.91	1.96	1.99	2.03	2.01
	9	1.99	2.05	2.09	2.16	2.18
	10	2.06	2.12	2.17	2.27	2.31
4	4	1.04	0.98	0.87		
	5	1.27	1.22	1.15		
	6	1.43	1.39	1.35		
	7	1.55	1.53	1.50	1.37	
	8	1.62	1.62	1.61	1.53	
	9	1.69	1.71	1.70	1.67	1.54
	10	1.75	1.77	1.78	1.77	1.69
5	4	0.87	0.78			
	5	1.07	0.99	0.88		
	6	1.22	1.16	1.08		
	7	1.33	1.29	1.23		
	8	1.43	1.39	1.35	1.17	
	9	1.49	1.47	1.44	1.32	
	10	1.55	1.54	1.52	1.42	1.21
6	4	0.73	0.62			
	5	0.92	0.81			
	6	1.05	0.97	0.87		
	7	1.17	1.09	1.02		
	8	1.26	1.20	1.13		
	9	1.34	1.29	1.23	1.03	
	10	1.40	1.36	1.32	1.16	

Parameter estimation

Next, the parameter estimation using the maximum likelihood method is addressed and a simulation is conducted to study the performance of the method. Let $(x_1, x_2 \dots x_n)$ be a random sample from a GWD with parameters β, σ, a , and λ , then the log-likelihood function from (1) is given by

$$\begin{aligned} \ell(\beta, \sigma, a, \lambda) &= \ell = \sum_{i=1}^n \ln(g(x_i)) \\ &= n(\ln \beta + \ln a - \ln \sigma - a \ln \lambda) + (a-1) \sum_{i=1}^n \ln x_i - \frac{1}{\sigma} \sum_{i=1}^n \left(\frac{x_i}{\lambda}\right)^a \\ &\quad - \left(1 + \frac{1}{\sigma}\right) \sum_{i=1}^n \ln \left\{1 - e^{-(x_i/\lambda)^a}\right\} - \beta \sum_{i=1}^n \left\{e^{(x_i/\lambda)^a} - 1\right\}^{-1/\sigma}. \end{aligned} \quad (6)$$

The first partial derivatives of (6) are

$$\frac{\partial \ell}{\partial \beta} = \frac{n}{\beta} - \sum_{i=1}^n \left\{e^{(x_i/\lambda)^a} - 1\right\}^{-1/\sigma}, \quad (7)$$

$$\begin{aligned} \frac{\partial \ell}{\partial \sigma} &= -\frac{n}{\sigma} + \frac{1}{\sigma^2} \sum_{i=1}^n (x_i / \lambda)^a + \frac{1}{\sigma^2} \sum_{i=1}^n \ln \left\{1 - e^{-(x_i/\lambda)^a}\right\} \\ &\quad - \frac{\beta}{\sigma^2} \sum_{i=1}^n \left\{e^{(x_i/\lambda)^a} - 1\right\}^{-1/\sigma} \ln \left\{e^{(x_i/\lambda)^a} - 1\right\}, \end{aligned} \quad (8)$$

$$\begin{aligned} \frac{\partial \ell}{\partial a} &= \frac{n}{a} - n \ln \lambda + \sum_{i=1}^n \ln x_i - \frac{1}{\sigma} \sum_{i=1}^n (x_i / \lambda)^a \ln(x_i / \lambda) \\ &\quad - (1 + 1/\sigma) \sum_{i=1}^n (x_i / \lambda)^a \ln(x_i / \lambda) \left\{e^{(x_i/\lambda)^a} - 1\right\}^{-1} \\ &\quad + \frac{\beta}{\sigma} \sum_{i=1}^n \left\{e^{(x_i/\lambda)^a} - 1\right\}^{-1-1/\sigma} e^{(x_i/\lambda)^a} (x_i / \lambda)^a \ln(x_i / \lambda), \end{aligned} \quad (9)$$

and

GUMBEL-WEIBULL DISTRIBUTION

$$\begin{aligned} \frac{\partial \ell}{\partial \lambda} = & -\frac{na}{\lambda} + \frac{a}{\lambda \sigma} \sum_{i=1}^n (x_i / \lambda)^a + \frac{a}{\lambda} (1 + 1/\sigma) \sum_{i=1}^n (x_i / \lambda)^a \left\{ e^{(x_i / \lambda)^a} - 1 \right\}^{-1} \\ & - \frac{a\beta}{\lambda \sigma} \sum_{i=1}^n (x_i / \lambda)^a e^{(x_i / \lambda)^a} \left\{ e^{(x_i / \lambda)^a} - 1 \right\}^{-1-1/\sigma}. \end{aligned} \quad (10)$$

On setting (7) through (10) equal to zero and solving the system of equations iteratively using NLMIXED procedure in SAS, the maximum likelihood estimates of the parameters are obtained.

Because the parameters a and λ are from Weibull distribution (WD), and the parameters β and σ are from Gumbel distribution (GD), moment estimates of WD parameters and GD parameters are used as initial estimates. By assuming that (x_1, x_2, \dots, x_n) has WD, $w_i = \ln(x_i)$ can be computed. The initial estimates for a and λ are $a_0 = \pi / (s_w \sqrt{6})$ and $\lambda_0 = \exp(\bar{w} - \gamma / a_0)$, where \bar{w} and s_w are respectively the mean and the standard deviation of the sample (w_1, \dots, w_n) (Johnson et al., 1994, pp. 635-643), and $\gamma = -\Gamma'(1) \approx 0.57722$ is Euler's constant. The random sample (x_1, x_2, \dots, x_n) is transformed to a sample from GD by $z_i = \ln(\exp(x_i / \lambda_0)^{a_0} - 1)$. The initial estimates for β and σ are $\beta_0 = e^{v_0/\sigma_0}$ and $\sigma_0 = s_z \sqrt{6} / \pi$, where $v_0 = \bar{z} - \gamma \sigma_0$ and σ_0 are the moment estimates from GD, \bar{z} and s_z are respectively the mean and standard deviation of (z_1, \dots, z_n) (Johnson et al., 1995, p. 12).

A simulation study is conducted to examine the performance of the maximum likelihood estimates (MLEs). The bias (actual – estimate) and standard deviation are used to measure the performance. Three sample sizes $n = 250, 500$ and 1000 are used. The data are simulated from standard exponential distribution and then transformed into GWD using the transformation $X = \lambda \left(\ln \{ (Y / \beta)^{-\sigma} + 1 \} \right)^{1/a}$. The process is repeated 100 times for each parameter combination and each sample size. This simulation study is conducted for many parameter combinations, but for brevity the results for the parameter combinations $\beta = 0.5, 1, 4$, $\sigma = 2$, $a = 1, 2, 4$, and $\lambda = 1$ are reported.

The average bias and the standard deviation of the MLEs are computed and presented in Table 2. From the table it is observed that the bias appears reasonable and in general the standard deviation decreases as n increases. These observations also hold for the other parameter combinations that are not reported in Table 2. The simulation study shows that the maximum likelihood method is appropriate for estimating the GWD parameters.

Table 2. Average bias and standard deviation for MLEs when $\sigma = 2$ and $\lambda = 1$

β	a	n	$\hat{\beta}$		$\hat{\sigma}$		\hat{a}		$\hat{\lambda}$	
			bias	sd	bias	sd	bias	sd	bias	sd
0.5	1	250	-0.0096	0.0867	-0.0354	0.1514	-0.0734	0.2750	-0.0307	0.2047
		500	0.0082	0.0830	-0.0473	0.1326	-0.0853	0.2192	-0.0574	0.1931
		1000	0.0097	0.0642	-0.0329	0.1003	-0.0532	0.1565	-0.0471	0.1482
	2	250	0.0065	0.0942	-0.1094	0.3163	-0.0893	0.2785	-0.0336	0.1140
		500	0.0136	0.0879	-0.1103	0.2745	-0.0921	0.2170	-0.0341	0.1054
		1000	0.0110	0.0652	-0.0674	0.2015	-0.0532	0.1565	-0.0230	0.0754
	4	250	0.0065	0.0942	-0.2188	0.6325	-0.0893	0.2785	-0.0151	0.0562
		500	0.0148	0.0874	-0.2270	0.5470	-0.0935	0.2162	-0.0156	0.0512
		1000	0.0114	0.0650	-0.1363	0.4031	-0.0539	0.1570	-0.0109	0.0372
1	1	250	-0.0049	0.1832	-0.0319	0.1281	-0.0644	0.2429	-0.0380	0.2055
		500	0.0443	0.1597	-0.0413	0.1030	-0.0482	0.1738	-0.0862	0.1872
		1000	0.0346	0.1375	-0.0300	0.0893	-0.0317	0.1248	-0.0625	0.1626
	2	250	0.0335	0.2084	-0.1343	0.3083	-0.0905	0.2539	-0.0462	0.1276
		500	0.0663	0.1723	-0.1317	0.2518	-0.0755	0.1812	-0.0550	0.1061
		1000	0.0426	0.1435	-0.0790	0.1956	-0.0431	0.1287	-0.0348	0.0845
	4	250	0.0335	0.2084	-0.2685	0.6166	-0.0905	0.2539	-0.0210	0.0625
		500	0.0682	0.1715	-0.2591	0.4872	-0.0722	0.1788	-0.0263	0.0512
		1000	0.0415	0.1431	-0.1498	0.3893	-0.0410	0.1296	-0.0160	0.0416
	1	250	-0.0876	0.7596	-0.0017	0.0956	-0.1131	0.4299	0.0160	0.2424
		500	0.0049	0.6856	-0.0160	0.0792	-0.0755	0.2637	-0.0192	0.2154
		1000	-0.0265	0.7518	-0.0021	0.0774	-0.0249	0.2223	-0.0195	0.2401
4	2	250	0.3277	0.8897	-0.1143	0.2423	-0.0298	0.3595	-0.0722	0.1654
		500	0.3395	0.8917	-0.1004	0.2081	0.0101	0.2951	-0.0761	0.1637
		1000	0.3088	0.8031	-0.0810	0.1858	0.0276	0.2184	-0.0657	0.1430
	4	250	0.3712	0.8546	-0.1404	0.4622	-0.0696	0.4325	-0.0225	0.0836
		500	0.3876	0.8298	-0.1720	0.3838	-0.0244	0.2715	-0.0283	0.0754
		1000	0.2294	0.8272	-0.1171	0.3705	-0.0106	0.2208	-0.0189	0.0720

Tests about parameter β

When $\beta = 1$, the GWD reduces to

$$g_0(x) = \frac{a}{\lambda \sigma} (x/\lambda)^{a-1} e^{(x/\lambda)^a} \left(e^{(x/\lambda)^a} - 1 \right)^{-1-1/\sigma} \exp \left\{ - \left(e^{(x/\lambda)^a} - 1 \right)^{-1/\sigma} \right\}. \quad (11)$$

The three-parameter GWD in (11) can be compared with the four-parameter GWD in (1) by testing the null hypothesis

$$H_0 : \beta = 1 \text{ against } H_1 : \beta \neq 1. \quad (12)$$

GUMBEL-WEIBULL DISTRIBUTION

The hypotheses in (12) are tested by using the likelihood ratio test (Neyman and Pearson, 1928), the Wald test (Wald, 1943), or the score test (Rao, 1948).

The likelihood ratio statistic for testing (12) is

$\eta = -2[\ell_0(\tilde{\sigma}, \tilde{a}, \tilde{\lambda}) - \ell_1(\hat{\beta}, \hat{\sigma}, \hat{a}, \hat{\lambda})]$, where $\ell_0(\tilde{\sigma}, \tilde{a}, \tilde{\lambda})$ is the log-likelihood value of the three-parameter GWD and $\ell_1(\hat{\beta}, \hat{\sigma}, \hat{a}, \hat{\lambda})$ is the log-likelihood value of the four-parameter GWD. The likelihood ratio statistic η has asymptotic chi-square distribution with one degree of freedom.

The Wald statistic for testing (12) is $Z_{\hat{\beta}} = (\hat{\beta} - 1) / SE_{\hat{\beta}}$, where $\hat{\beta}$ is the MLE from the four-parameter GWD and $SE_{\hat{\beta}}$ is the standard error of $\hat{\beta}$. The Wald statistic $Z_{\hat{\beta}}$ has an approximate standard normal distribution.

The score statistic for testing (12) is $S = V^T I^{-1} V$, where the score vector V is the 4×1 gradient vector of ℓ with entries $V_i = \partial \ell / \partial \theta_i$, I is the 4×4 information matrix with entries $I_{i,j} = -E[\partial^2 \ell / \partial \theta_i \partial \theta_j]$, ℓ is the GWD log-likelihood function in (6), $\theta_1 = \beta$, $\theta_2 = \sigma$, $\theta_3 = a$, and $\theta_4 = \lambda$, and S is computed under the null hypothesis $\beta = 1$. The score statistic S has an approximate chi-square distribution with one degree of freedom.

A simulation study is conducted to compare the powers of the three tests for parameter β . Three sample sizes $n = 250, 500$ and 1000 are used in the study. The data are simulated from GWD as described under Parameter Estimation, and the parameters are estimated by the method of maximum likelihood. The simulation is repeated 200 times for each parameter combination and each sample size. The proportion of times of rejecting H_0 is used to estimate the power of each test. Two significance levels, 5% and 10%, are used and the results are similar. Many parameter combinations are used in the simulation. For brevity, Table 3 reports the results for 5% and the parameter values $\beta = 0.5, 1, 2, 4$, $\sigma = 1.6, 2$, $a = 1, 2, 3, 4$, and $\lambda = 1$.

When $\beta \leq 1$, it is observed from Table 3 that the score test is the most powerful followed by the Wald test while the likelihood ratio test is poor. On the contrary, when $\beta > 1$, the likelihood ratio test is the most powerful followed by the score test and the Wald test. In general, the power increases as the sample size increases.

Table 3. The proportion of times the null hypothesis is rejected when $\lambda = 1$

σ	β	a	$n=250$			$n=500$			$n=1000$		
			LR	Wald	Score	LR	Wald	Score	LR	Wald	Score
1.6	0.5	1	0.000	0.005	0.430	0.000	0.045	0.745	0.020	0.600	0.955
		2	0.000	0.000	0.420	0.000	0.060	0.685	0.040	0.600	0.940
		3	0.000	0.005	0.435	0.000	0.060	0.650	0.010	0.530	0.945
		4	0.000	0.000	0.495	0.000	0.060	0.690	0.020	0.620	0.970
	1	1	0.000	0.000	0.015	0.000	0.005	0.050	0.005	0.040	0.070
		2	0.000	0.000	0.060	0.000	0.060	0.165	0.015	0.125	0.145
		3	0.000	0.000	0.070	0.000	0.060	0.140	0.010	0.060	0.095
		4	0.000	0.000	0.070	0.000	0.045	0.120	0.015	0.055	0.080
	2	1	0.090	0.000	0.005	0.340	0.010	0.020	0.740	0.110	0.405
		2	0.070	0.000	0.005	0.220	0.000	0.020	0.720	0.055	0.340
		3	0.070	0.000	0.000	0.315	0.005	0.050	0.670	0.095	0.395
		4	0.065	0.000	0.005	0.295	0.000	0.030	0.675	0.115	0.350
	4	1	0.605	0.010	0.120	0.980	0.050	0.450	1.000	0.515	0.975
		2	0.585	0.015	0.075	0.935	0.040	0.435	1.000	0.400	0.945
		3	0.490	0.000	0.065	0.935	0.025	0.385	1.000	0.410	0.945
		4	0.650	0.030	0.100	0.935	0.075	0.475	1.000	0.585	0.980
2	0.5	1	0.000	0.090	0.660	0.030	0.560	0.945	0.480	0.970	1.000
		2	0.000	0.085	0.715	0.020	0.495	0.925	0.545	0.990	1.000
		3	0.000	0.080	0.740	0.035	0.490	0.940	0.525	0.985	1.000
		4	0.005	0.120	0.745	0.025	0.575	0.960	0.545	0.985	1.000
	1	1	0.000	0.000	0.050	0.000	0.000	0.060	0.005	0.030	0.090
		2	0.000	0.020	0.100	0.010	0.050	0.110	0.030	0.070	0.080
		3	0.000	0.000	0.085	0.000	0.065	0.105	0.025	0.045	0.055
		4	0.000	0.010	0.090	0.005	0.065	0.080	0.010	0.025	0.030
	2	1	0.360	0.005	0.030	0.730	0.080	0.360	0.985	0.760	0.925
		2	0.345	0.010	0.060	0.655	0.130	0.345	0.920	0.715	0.820
		3	0.290	0.005	0.055	0.645	0.105	0.360	0.940	0.735	0.830
		4	0.290	0.005	0.040	0.715	0.115	0.370	0.930	0.775	0.860
	4	1	0.940	0.160	0.425	1.000	0.515	0.935	1.000	1.000	1.000
		2	0.840	0.110	0.285	1.000	0.380	0.835	1.000	0.955	1.000
		3	0.875	0.100	0.305	1.000	0.485	0.885	1.000	0.940	1.000
		4	0.880	0.095	0.365	1.000	0.510	0.905	1.000	0.965	1.000

Applications

Applications of the GWD to five data sets are now presented. These five data sets exhibit various shapes of distribution including right-skewed, approximately symmetric, left-skewed, reversed J-shape and bimodal distributions for demonstrating the flexibility of the GWD for fitting real world data. The parameters are estimated by using the method of maximum likelihood. The fit is compared to

GUMBEL-WEIBULL DISTRIBUTION

other distributions based on the p-value of the Kolmogorov-Smirnov (K-S) statistic, and the Akaike information criterion (AIC).

Breaking stress of carbon fibers data

The data for breaking stress of carbon fibers of 50 mm length (GPa) in Table 4 is obtained from Nicholas and Padgett (2006). This data was used by Cordeiro and Lemonte (2011) to illustrate the application of the four-parameter beta-Birnbaum-Saunders distribution (BBS) when compared to the two-parameter Birnbaum-Saunders distribution (Birnbaum and Saunders, 1969). The data set is unimodal and is approximately symmetric (skewness = -0.13 and kurtosis = 0.34).

Table 4. Breaking stress of carbon fibers data

0.39	0.85	1.08	1.25	1.47	1.57	1.61	1.61	1.69	1.80	1.84
1.87	1.89	2.03	2.03	2.05	2.12	2.35	2.41	2.43	2.48	2.50
2.53	2.55	2.55	2.56	2.59	2.67	2.73	2.74	2.79	2.81	2.82
2.85	2.87	2.88	2.93	2.95	2.96	2.97	3.09	3.11	3.11	3.15
3.15	3.19	3.22	3.22	3.27	3.28	3.31	3.31	3.33	3.39	3.39
3.56	3.60	3.65	3.68	3.70	3.75	4.20	4.38	4.42	4.70	4.90

Table 5. MLEs for breaking stress of carbon fibers data (standard errors in parentheses)

Distribution	BBS*	BED	BGE	GWD
Parameter estimates	$\hat{a} = 0.1930$ (0.0259)	$\hat{a} = 7.5072$ (0.7642)	$\hat{a} = 0.6473$ (0.3077)	$\hat{\beta} = 3.4359$ (1.1494)
	$\hat{b} = 1876.7324$ (605.05)	$\hat{b} = 20.9967$ (1.4865)	$\hat{b} = 1198.50$ (5.9057)	$\hat{\sigma} = 5.5673$ (2.8064)
	$\hat{\alpha} = 1.0445$ (0.0036)	$\hat{\lambda} = 0.1131$ (0.0170)	$\hat{\alpha} = 5.099$ (1.9670)	$\hat{a} = 2.4231$ (0.5078)
	$\hat{\beta} = 57.6001$ (0.3313)		$\hat{\lambda} = 0.0824$ (0.0419)	$\hat{\lambda} = 1.1324$ (0.4524)
Log Likelihood	- 91.36	- 91.22	- 85.96	- 84.83
AIC	190.7	188.4	179.9	177.7
K-S	0.1422	0.1338	0.0817	0.0666
p-value	0.1384	0.1884	0.7710	0.9313

*Parameter estimates and value of AIC statistic from Cordeiro and Lemonte (2011).

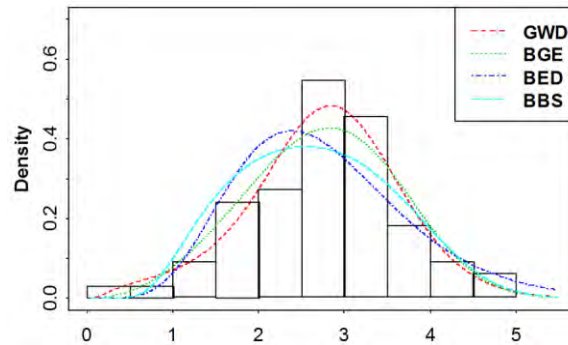


Figure 5. Histogram and the fitted PDFs for the breaking stress of carbon fibers data

Four distributions are used to fit the data: the beta-exponential (BED) defined by Nadarajah and Kotz (2006), the beta-generalized exponential (BGE) defined by Barreto-Souza et al. (2010), the beta-Birnbaum-Saunders, and GWD. The MLEs and goodness of fit statistics are presented in Table 5. The MLEs, standard errors of MLEs, and the AIC statistic of BBS distribution are obtained from Cordeiro and Lemonte (2011). Figure 5 displays the histogram and the PDFs of the fitted distributions.

To compare the four-parameter GWD in (1) to the three-parameter GWD in (11), the likelihood ratio statistic $\eta = 14.302$, the Wald statistic $Z_{\hat{\beta}} = 2.119$, and the score statistic $S = 6.630$ are obtained. All the test statistics lead to rejecting the null hypothesis of three-parameter GWD in favor of the four-parameter GWD at 5% significance level. The four-parameter GWD outperforms the other three distributions by comparing the AIC and K-S statistics. This application suggests that the four-parameter GWD fits unimodal symmetric data very well when compared to the other distributions.

Australian athletes' data

The Australian athletes' data reported by Cook and Weisberg (1994) contains 13 variables on 102 male and 100 female Australian athletes collected at the Australian Institute of Sport. Jamalizadeh et al. (2011) used the heights for the 100 female athletes and the hemoglobin concentration levels for the 202 athletes to illustrate the application of a generalized skew two-piece skew-normal distribution. Choudhury and Abdul Matin (2011) also used percentage of the hemoglobin blood cell for the male athletes to illustrate the application of an extended skew

GUMBEL-WEIBULL DISTRIBUTION

generalized normal distribution. Next, the GWD was applied to two of the variables in the Australian athletes' data, the sum of skin folds (SSF) and the height in centimeters for the 100 female athletes.

The Weibull-Pareto distribution (WPD) defined by Alzaatreh, Famoye and Lee (2013), the beta-normal distribution (BND) defined by Eugene et al. (2002), the exponentiated-Weibull (EWD) defined by Mudholkar and Srivastava (1993), and the GWD are used to fit both data sets.

The sum of skin folds for the 100 female athletes in Table 6 is right skewed (skewness = 0.79, kurtosis = 0.73). The MLEs and the goodness of fit statistics are in Table 7. The histogram and the densities of the fitted distributions are provided in Figure 6.

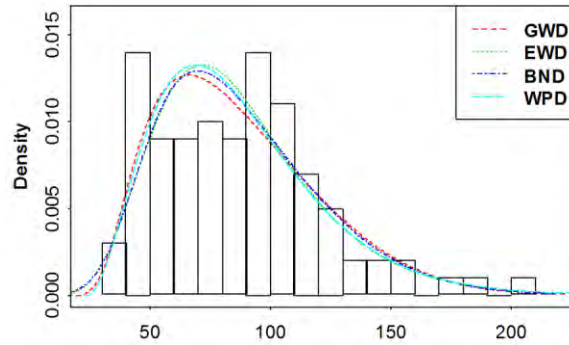
Table 6. The sum of skin folds data

33.8	36.8	38.2	41.1	41.6	42.3	43.5	43.5	46.1	46.2	46.3	47.5
47.6	48.4	49.0	49.9	50.0	52.5	52.6	54.6	54.6	55.6	56.8	57.9
58.9	59.4	61.9	62.6	62.9	65.1	67.0	68.3	68.9	69.9	70.0	71.3
71.6	73.9	74.7	74.9	75.1	75.2	76.2	76.8	77.0	80.1	80.3	80.3
80.3	80.6	83.0	87.2	88.2	89.0	90.2	90.4	91.0	91.2	95.4	96.8
97.2	97.9	98.0	98.1	98.3	98.5	99.8	99.9	101.1	102.8	102.8	103.6
103.6	104.6	106.9	109.0	109.1	109.5	109.6	110.2	110.7	111.1	113.5	114.0
115.9	117.8	122.1	123.6	125.9	126.4	126.4	131.9	136.3	143.5	148.9	156.6
156.6	171.1	181.7	200.8								

The three tests about parameter β are used to compare the three- and four-parameter GWD and the tests are not significant. Thus, the three-parameter GWD is adequate to fit the data. On fitting the three-parameter GWD, the K-S statistic shows that it is not as good as the four-parameter GWD. By comparing the goodness of fit statistics among the four distributions in Table 7, it is observed that all distributions provide good fits. The GWD provides the best fit based on the K-S statistic. This application suggests that GWD fits right skewed data very well.

Table 7. MLEs for sum of skin folds data (standard errors in parentheses)

Distribution	WPD	BND	EWD	GWD
Parameter estimates	$\hat{c} = 3.6308$ (1.0568)	$\hat{a} = 9.7706$ (0.4008)	$\hat{a} = 1.2245$ (0.4378)	$\hat{\beta} = 1.0299$ (1.0094)
	$\hat{\beta} = 0.7183$ (0.1799)	$\hat{b} = 0.1967$ (0.0223)	$\hat{\theta} = 6.5038$ (6.3413)	$\hat{\sigma} = 1.3303$ (0.3218)
	$\hat{\theta} = 23.0497$ (7.6212)	$\hat{\mu} = 9.1517$ (4.3612)	$\hat{\sigma} = 41.4980$ (24.4832)	$\hat{a} = 2.2896$ (0.9184)
		$\hat{\sigma} = 25.4309$ (0.9049)		$\hat{\lambda} = 81.7841$ (27.2786)
Log Likelihood	- 486.07	- 487.06	- 487.17	- 485.89
AIC	978.1	982.1	980.3	979.8
K-S	0.0825	0.0711	0.0808	0.0705
p-value	0.5043	0.6925	0.5307	0.7022

**Figure 6.** Histogram and the fitted PDFs for sum of skin folds data

The height in centimeters for the 100 Australian female athletes in Table 8 is unimodal and left skewed (skewness = -0.57 , kurtosis = 1.32). The MLEs and the goodness of fit statistics are in Table 9. The histogram and the PDFs of the fitted distributions are plotted in Figure 7.

The hypothesis in (12) is used to test the significance of the parameter β . All the tests, the likelihood ratio ($\eta = 13.772$), the Wald ($Z_{\hat{\beta}} = 2.309$), and the score

GUMBEL-WEIBULL DISTRIBUTION

($S = 4.487$) reject the null hypothesis in favor of the four-parameter GWD. By comparing the goodness of fit statistics among the four distributions it was found that all distributions are competitors as they provide good fits to the data. The four-parameter GWD appears to provide the best p-value of the K-S statistic, and shares the best value of the AIC statistic with the exponentiated-Weibull distribution. This application suggests that GWD fits left skewed unimodal data very well.

Table 8. The heights data

148.9	149.0	156.0	156.9	157.9	158.9	162.0	162.0	162.5	163.0	163.9	165.0
166.1	166.7	167.3	167.9	168.0	168.6	169.1	169.8	169.9	170.0	170.0	170.3
170.8	171.1	171.4	171.4	171.6	171.7	172.0	172.2	172.3	172.5	172.6	172.7
173.0	173.3	173.3	173.5	173.6	173.7	173.8	174.0	174.0	174.0	174.1	174.1
174.4	175.0	175.0	175.0	175.3	175.6	176.0	176.0	176.0	176.0	176.8	177.0
177.3	177.3	177.5	177.5	177.8	177.9	178.0	178.2	178.7	178.9	179.3	179.5
179.6	179.6	179.7	179.7	179.8	179.9	180.2	180.2	180.5	180.5	180.9	181.0
181.3	182.1	182.7	183.0	183.3	183.3	184.6	184.7	185.0	185.2	186.2	186.3
188.7	189.7	193.4	195.9								

Table 9. MLEs for heights data (standard errors in parentheses)

Distribution	WPD	BND	EWD	GWD
Parameter estimates	$\hat{c} = 8.1823$ (3.3244)	$\hat{a} = 0.8697$ (0.6607)	$\hat{a} = 14.7405$ (3.2361)	$\hat{\beta} = 6.2028$ (2.2535)
	$\hat{\beta} = 2.8449$ (1.1079)	$\hat{b} = 6.3120$ (0.7454)	$\hat{\theta} = 2.7836$ (1.3390)	$\hat{\sigma} = 3.9326$ (2.9597)
	$\hat{\theta} = 125.1377$ (16.9742)	$\hat{\mu} = 191.46$ (3.9729)	$\hat{\sigma} = 170.26$ (4.5462)	$\hat{a} = 12.268$ (2.1530)
		$\hat{\sigma} = 12.0376$ (3.1219)		$\hat{\lambda} = 147.02$ (12.5835)
Log Likelihood	- 351.03	- 350.31	- 350.37	- 349.33
AIC	708.1	708.6	706.7	706.7
K-S	0.0801	0.0725	0.0711	0.0565
p-value	0.5427	0.6692	0.6923	0.9071

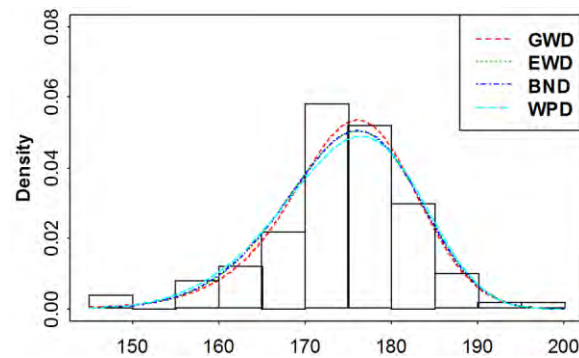


Figure 7. Histogram and the fitted PDFs for heights data

Application to lifetime data

Two data sets that represent the stress rupture life in hours of Kevlar 49/epoxy strands when subjected to a constant sustained stress level pressure until failure are used to illustrate an application to lifetime data. Both data sets are from Andrews and Herzberg (1985) and the original source is Barlow et al. (1984). Cooray and Ananda (2008) used the data sets to illustrate the usefulness of the generalized half normal distribution (GHN) when compared to other commonly used distributions. Four distributions are used to fit both data sets: the exponentiated-Weibull, beta-normal, generalized half normal, and GWD. The MLEs, log-likelihood value, and the value of K-S statistic of GHN distribution are obtained from Cooray and Ananda (2008).

The data set in Table 10 represents the failure times of Kevlar 49/epoxy strands when the pressure is at 90% stress level. This data is leptokurtic, unimodal, highly right skewed, and has reversed-J shape with a potential outlier (skewness = 3.05, kurtosis = 14.47).

The MLEs, log likelihood values, values of AIC and K-S, and p-values of the K-S are in Table 11. Figure 8 contains the histogram and the PDFs of the fitted distributions. The three tests about parameter β are used to test its significance. Only the likelihood ratio test rejects the null hypothesis of the three-parameter GWD, while the other two tests fail to reject the null hypothesis in (12). Because the likelihood ratio test is the most powerful when $\beta > 1$, one can conclude that the parameter β is significant.

GUMBEL-WEIBULL DISTRIBUTION

Table 10. Kevlar 49/epoxy strands failure times data (pressure at 90%)

0.01	0.01	0.02	0.02	0.02	0.03	0.03	0.04	0.05	0.06	0.07
0.07	0.08	0.09	0.09	0.10	0.10	0.11	0.11	0.12	0.13	0.18
0.19	0.20	0.23	0.24	0.24	0.29	0.34	0.35	0.36	0.38	0.40
0.42	0.43	0.52	0.54	0.56	0.60	0.60	0.63	0.65	0.67	0.68
0.72	0.72	0.72	0.73	0.79	0.79	0.80	0.80	0.83	0.85	0.90
0.92	0.95	0.99	1.00	1.01	1.02	1.03	1.05	1.10	1.10	1.11
1.15	1.18	1.20	1.29	1.31	1.33	1.34	1.40	1.43	1.45	1.50
1.51	1.52	1.53	1.54	1.54	1.55	1.58	1.60	1.63	1.64	1.80
1.80	1.81	2.02	2.05	2.14	2.17	2.33	3.03	3.03	3.34	4.20
4.69	7.89									

Table 11. MLEs for Kevlar 49/epoxy strands failure times data (pressure at 90%) (standard errors in parentheses)

Distribution	GHN*	EWD	BND	GWD
Parameter estimates	$\hat{\theta} = 1.2238$	$\hat{a} = 1.0604$ (0.2399)	$\hat{a} = 12.4298$ (0.3071)	$\hat{\beta} = 1.8064$ (0.5037)
	$\hat{\alpha} = 0.7108$	$\hat{\theta} = 0.7929$ (0.2870)	$\hat{b} = 0.4467$ (0.0475)	$\hat{\sigma} = 3.2713$ (0.6459)
		$\hat{\sigma} = 1.2180$ (0.3933)	$\hat{\mu} = -1.5065$ (0.1347)	$\hat{a} = 0.9200$ (0.1594)
			$\hat{\sigma} = 1.1413$ (0.0317)	$\hat{\lambda} = 0.2071$ (0.1072)
Log Likelihood	- 103.33	- 102.79	- 132.98	- 100.23
AIC	210.7	211.6	274.0	208.5
K-S	0.0800	0.0844	0.1129	0.0687
p-value	0.5377	0.4680	0.1526	0.7266

*MLEs, log likelihood, and K-S from Cooray and Ananda (2008), the standard errors were not given.

By comparing the p-value of the K-S and the AIC among the four distributions, it is observed that the GWD outperforms the other distributions. This application suggests that the GWD fits unimodal reversed-J shape data very well.

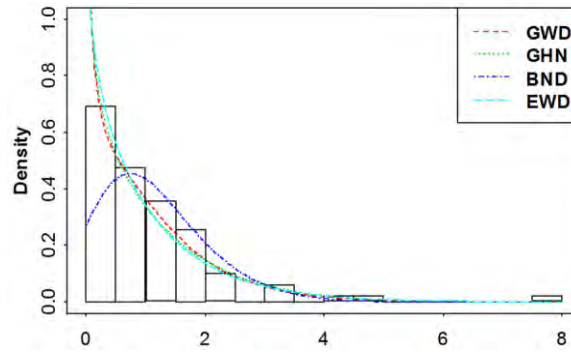


Figure 8. Histogram and the fitted PDFs for failure times data (90% pressure)

The second part of the Kevlar data in Table 12 represent the failure times when the pressure is at 70% stress level. This data is multimodal, platykurtic, and approximately symmetric (skewness = 0.1, kurtosis = -0.79) with very wide range. Table 13 contains the MLEs, AIC and K-S statistics, and the p-value of the K-S statistic. Figure 9 provides the histogram and the PDFs of the fitted distributions.

Table 12. Kevlar 49/epoxy strands failure times data (pressure at 70%)

1051	1337	1389	1921	1942	2322	3629	4006	4012	4063
4921	5445	5620	5817	5905	5956	6068	6121	6473	7501
7886	8108	8546	8666	8831	9106	9711	9806	10205	10396
10861	11026	11214	11362	11604	11608	11745	11762	11895	12044
13520	13670	14110	14496	15395	16179	17092	17568	17568	

When the three tests about β are used to compare the three-parameter GWD with the four-parameter GWD, the parameter β is not significantly different from 1. The three-parameter GWD was also fitted to the data and the result is comparable to that of four-parameter GWD. By using the K-S statistic, the three-parameter GWD is not as good as the four-parameter GWD.

By comparing the goodness of fit statistics in Table 13 among the four distributions, it was found that the four distributions are competitors and fit the data very well. The GWD and the beta-normal distribution capture the data with a bimodal density curve. The GHN with the smallest AIC and largest p-value of the K-S statistic does not capture the bimodality property. This application suggests that GWD is capable of providing an adequate fit to bimodal data.

GUMBEL-WEIBULL DISTRIBUTION

Table 13. MLEs for Kevlar 49/epoxy strands failure times data (pressure at 70%) (standard errors in parentheses)

Distribution	GHN*	EWD	BND	GWD
Parameter estimates	$\hat{\theta} = 10906.98$	$\hat{a} = 5.2226$ (2.9883)	$\hat{a} = 0.1626$ (0.1039)	$\hat{\beta} = 1.3118$ (0.5144)
	$\hat{\alpha} = 1.64067$	$\hat{\theta} = 0.2644$ (0.1887)	$\hat{b} = 0.1157$ (0.0199)	$\hat{\sigma} = 4.1091$ (1.0456)
		$\hat{\sigma} = 14559$ (2007.54)	$\hat{\mu} = 7826$ (1759.97)	$\hat{a} = 2.6948$ (0.8101)
			$\hat{\sigma} = 1339.35$ (245.62)	$\hat{\lambda} = 6165.69$ (1749.51)
Log Likelihood	- 479.56	- 479.03	- 480.52	- 478.51
AIC	963.1	964.1	969.0	965.0
K-S	0.067	0.0825	0.0797	0.0749
p-value	0.9804	0.8926	0.9144	0.9462

* MLEs, log likelihood, and K-S from Cooray and Ananda (2008), the standard errors were not given.

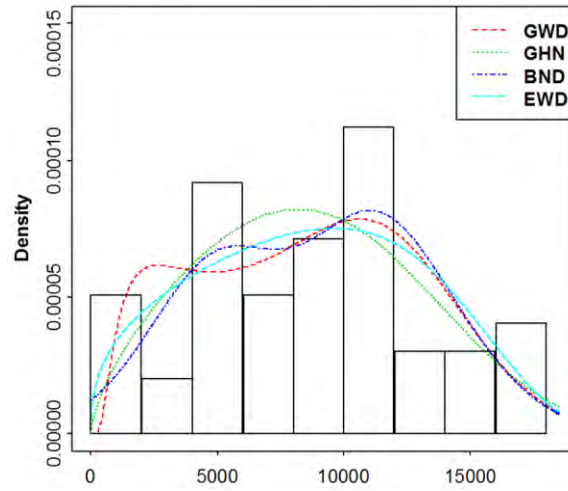


Figure 9. Histogram and the fitted PDFs for failure times data (70% pressure)

Summary and conclusion

In this article, some properties of the four-parameter GWD, a member of the $T-X$ family of distributions are provided. The paper provides important properties such as modes and mean deviations for GWD. The method of maximum likelihood is proposed for parameter estimation. The result of the simulation on the performance of the maximum likelihood method shows that the method is appropriate for estimating the parameters of GWD. Three test statistics are used to test the significance of the parameter β . The score test is the most powerful when $\beta \leq 1$, while the likelihood ratio test is the most powerful when $\beta > 1$. Five datasets are used to illustrate the application of the GWD and the results are compared to other existing distributions. The results of these applications suggest that GWD can be used to fit right-skewed, left-skewed, symmetric, unimodal and bimodal data.

References

- Al-Aqtash, R., Famoye, F. and Lee, C. (2014). On generating a new family of distributions using the logit function. To appear in the *Journal of Probability and Statistical Science*.
- Alzaatreh, A., Famoye, F. & Lee, C. (2013). Weibull-Pareto distribution and its applications, *Communications in Statistics – Theory and Methods*, 42, 1673-1691.
- Alzaatreh, A., Lee, C. & Famoye, F. (2013). A new method for generating families of continuous distributions, *Metron*, 71(1), 63-79.
- Andrews, D. F. & Herzberg, A. M. (1985). *Data: A Collection of Problems from Many Fields for the Student and Research Worker* (Springer Series in Statistics). New York: Springer.
- Barlow, R. E., Toland, R. H. & Freeman, T. (1984). A Bayesian analysis of stress-rupture life of Kevlar 49/epoxy spherical pressure vessels. In T. D. Dwivedi (Ed.), *Proceedings of the Canadian Conference in Applied Statistics*. New York: Marcel Dekker.
- Barreto-Souza, W., Santos, A. H. S. & Cordeiro, G. M. (2010). The beta generalized exponential distribution, *Journal of Statistical Computation and Simulation*, 80(2), 159-172.
- Beirlant, J., Goegebeur, Y., Segers, J. & Teugels, J. (2006). *Statistics of Extremes: Theory and Applications*. West Sussex, England: John Wiley and Sons Ltd.

GUMBEL-WEIBULL DISTRIBUTION

- Birnbaum, Z. W. & Saunders, S. C. (1969). A new family of life distributions. *Journal of Applied Probability*, 6, 319-327.
- Choudhury, K. & Abdul Matin, M. (2011). Extended skew generalized normal distribution. *Metron*, 69(3), 265-278.
- Cook, R. D. & Weisberg, S. (1994). *An Introduction to Regression Graphics*. New York: John Wiley and Sons, Inc.
- Cooray, K. & Ananda, M. (2008). A generalization of the half-normal distribution with applications to lifetime data. *Communications in Statistics – Theory and Methods*, 37, 1323-1337.
- Cordeiro, G. M. & Lemonte, A. J. (2011). The β -Birnbbaum-Saunders distribution: An improved distribution for fatigue life modeling. *Computational Statistics and Data Analysis*, 55(3), 1445-1461.
- Eugene, N., Lee, C. & Famoye, F. (2002). Beta-normal distribution and its applications. *Communications in Statistics – Theory and Methods*, 31(4), 497-512.
- Gumbel, E. J. (1958). *Statistics of Extremes*. New York: Columbia University Press.
- Jamalizadeh, A., Arabpour, A.R., & Balakrishnan, N. (2011). A generalized skew two-piece skew-normal distribution. *Statistical Papers*, 52(2), 431-446.
- Johnson, N. L., Kotz, S. & Balakrishnan, N. (1994). *Continuous Univariate Distributions, Vol. 1* (2nd ed.). New York: John Wiley and Sons, Inc.
- Johnson, N. L., Kotz, S. & Balakrishnan, N. (1995). *Continuous Univariate Distributions, Vol. 2* (2nd ed.). New York: John Wiley and Sons, Inc.
- Kotz, S. & Nadarajah, S. (2000). *Extreme Value Distributions*. London: Imperial College Press.
- Mudholkar, G. S. & Srivastava, D. K. (1993). Exponentiated Weibull family for analyzing bathtub failure-rate data. *IEEE Transactions on Reliability*, 42(2), 299-302.
- Nadarajah, S. & Kotz, S. (2006). The beta exponential distribution, *Reliability Engineering and System Safety*, 91(6), 689-697.
- Neyman, J. & Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika*, 20, 175-240, 263-294.
- Nicholas, M. D. & Padgett, W. J. (2006). A bootstrap control for Weibull percentiles. *Quality and Reliability Engineering International*, 22, 141-151.

Rao. C. R. (1948). Large sample tests of statistical hypotheses concerning several parameters with application to problems of estimation. *Proceedings of the Cambridge Philosophical Society*, 44, 50-57.

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54(3), 426-482.

Objective Priors for Estimation of Extended Exponential Geometric Distribution

Pedro L. Ramos

Universidade de São Paulo
São Paulo, Brazil

Fernando A. Moala

Universidade Estadual Paulista
São Paulo, Brazil

Jorge A. Achcar

Universidade de São Paulo
São Paulo, Brazil

A Bayesian analysis was developed with different noninformative prior distributions such as Jeffreys, Maximal Data Information, and Reference. The aim was to investigate the effects of each prior distribution on the posterior estimates of the parameters of the extended exponential geometric distribution, based on simulated data and a real application.

Keywords: Extended exponential geometric distribution, Jeffreys, MDIP, Reference, Bayesian, noninformative, prior

Introduction

Adamidis & Loukas (2005) introduced an extension of the exponential geometric distribution (Adamidis & Loukas, 1998), naming it as an extended exponential geometric (EEG) distribution, to analyze lifetime data. This distribution provides increasing or decreasing hazard functions, depending on the values of its parameters. In this way, EEG gives a great flexibility of fit for the data.

If T is a random variable denoting the lifetime of a component with an extended exponential geometric (EEG) distribution, then the probability density is given by:

$$f(t|\gamma, \lambda) = \frac{\lambda \gamma e^{-\lambda t}}{(1 - (1 - \gamma)e^{-\lambda t})^2}, \quad (1)$$

with $t > 0$ and parameters $\gamma > 0$ and $\lambda > 0$. Let us denote this distribution as EEG(γ, λ).

Dr. Ramos is in the Faculty of Medicine of Ribeirão Preto. Email him at: pedrolramos@hotmail.com. Dr. Moala is a Professor in the Faculty of Science and Technology. Email him at: femoala@fct.unesp.br. Dr. Achcar is a Professor in the Institute of Mathematics and Computer Science. Email him at: achcar@fmrp.usp.br.

The survival and hazard functions of EEG(γ, λ) distribution, for a fixed time t , is given by

$$S(t; \gamma, \lambda) = \frac{\gamma e^{-\lambda t}}{1 - (1 - \gamma) e^{-\lambda t}} \text{ and } h(t; \gamma, \lambda) = \frac{\lambda}{1 - (1 - \gamma) e^{-\lambda t}}, \quad (2)$$

respectively.

The mean and variance of the EEG distribution are given, respectively, by

$$\begin{aligned} E(T) &= \frac{\gamma}{\lambda} \Psi(1 - \gamma, 1, 1) \text{ and} \\ \text{var}(T) &= \frac{\gamma}{\lambda^2} (2\Psi(1 - \gamma, 2, 1) - \gamma \Psi^2(1 - \gamma, 1, 1)) \end{aligned} \quad (3)$$

where $\Psi(z, s, a)$ is known as Lerch transcendental function (Erdelyi et al., 1953), given by

$$\Psi(z, s, a) = \frac{1}{\Gamma(s)} \int_0^\infty \frac{u^{s-1} e^{-au}}{1 - ze^{-u}} du = \sum_{k=0}^\infty \frac{z^k}{(a+k)^s}; z < 1; a, s > 0.$$

Adamidis et al. (2005) and Kitidamrongsuk (2010) gave additional properties of the EEG distribution.

Figures 1 and 2 present different forms for the density, survival and hazard functions for the EEG distribution considering different values of γ and λ .

The motivation here is to present a Bayesian analysis when there is little prior knowledge available or that reflects mainly the information from the sample.. In this situation, it is important to use noninformative priors, however, it can be difficult to choose a prior distribution that represent one of this situations. Thus, the main aim of this paper is to choose a noninformative prior distribution is for the parameters parameters λ and γ of the EEG distribution and to study the effects of these different priors in the resulting posterior distributions, especially in situations of small sample sizes, a common situation in applications.

OBJECTIVE PRIORS FOR ESTIMATION OF GEOMETRIC DISTRIBUTION

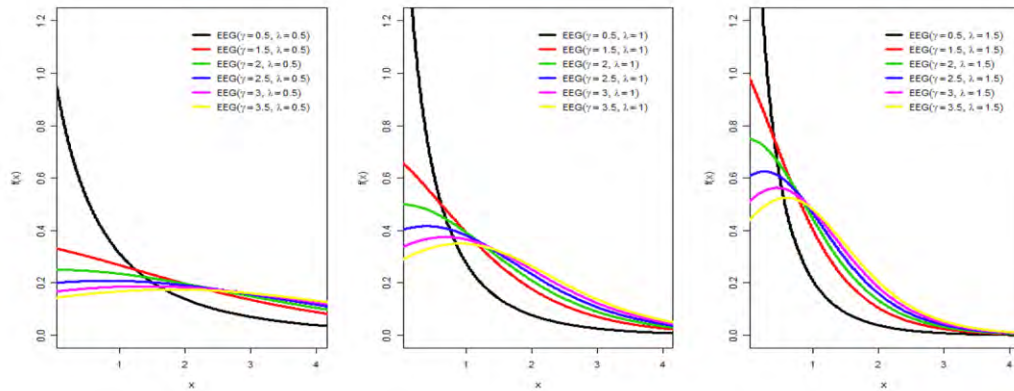


Figure 1. Probability density functions for the EEG distribution with values for the scale and shape parameters given respectively, by $\lambda = 0.5, 1.0, 1.5$ and $\gamma = 0.5, 1.5, 2.0, 2.5, 3.0, 3.5$.

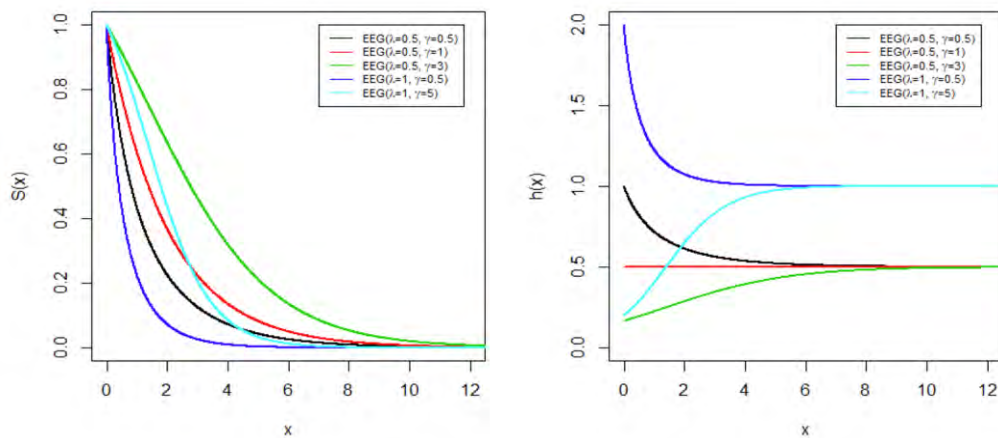


Figure 2. Survival functions and hazard functions for the EEG distribution with values for the scale and shape parameters given respectively, by $\lambda = 0.5, 1.0$ and $\gamma = 0.5, 1, 3, 5$.

Commonly used noninformative prior distributions are derived, such as uniform (Bayes, 1763; Laplace, 1774), Jeffreys (1967), reference priors (Bernardo, 1979; Berger & Bernardo, 1992), and the uncommon MDIP prior (Zellner, 1977, 1984). A simulation study is conducted comparing their performance in terms of their summaries and coverage rates of credible intervals.

Numerical integration based on stochastic simulation methods as the Markov Chain Monte Carlo (MCMC) will be used to simulate samples of the marginal posterior distribution of interest. In particular, we will be using the Metropolis-Hastings algorithm to obtain the posterior summaries of interest (see [Gelfand & Smith, 1990](#) or [Chib & Greenberg, 1995](#)).

Methodology

Maximum Likelihood Estimation

Let X_1, \dots, X_n be a random sample from $EEG(\gamma, \lambda)$ distribution with density (1). The likelihood function for the parameters γ and λ is given by

$$L(x; \gamma, \lambda) = (\lambda\gamma)^n \exp\left(-\lambda \sum_{i=1}^n x_i\right) \prod_{i=1}^n \left\{ \left(1 - (1-\gamma)e^{-\lambda x_i}\right)^{-2} \right\}, \quad (4)$$

where $\gamma > 0$ and $\lambda > 0$.

The logarithm of the likelihood function (4) is given by

$$l(x; \gamma, \lambda) = n \log(\lambda\gamma) - \lambda \sum_{i=1}^n x_i - 2 \sum_{i=1}^n \log\left(1 - (1-\gamma)e^{-\lambda x_i}\right). \quad (5)$$

By setting $\partial l(x; \gamma, \lambda) / \partial \gamma = 0$ and $\partial l(x; \gamma, \lambda) / \partial \lambda = 0$ and after some algebraic manipulations, we obtain the likelihood equations

$$\begin{aligned} \frac{n}{\lambda} - \sum_{i=1}^n x_i - 2(1-\gamma) \sum_{i=1}^n \frac{x_i e^{-\lambda x_i}}{1 - (1-\gamma)e^{-\lambda x_i}} &= 0 \text{ and} \\ \frac{n}{\lambda} - 2 \sum_{i=1}^n \frac{e^{-\lambda x_i}}{1 - (1-\gamma)e^{-\lambda x_i}} &= 0 \end{aligned} \quad (6)$$

whose solutions provide the maximum likelihood estimators of the parameters γ and λ . Note that the solutions of the likelihood equations (6) cannot be obtained analytically and hence numerical approaches need to be used.

Adamidis et al. (2005) propose to use the EM algorithm ([Dempster et al., 1977](#)) to solve the nonlinear equations (6) and find the MLE of γ and λ . The EM iterations are given by

$$\begin{aligned}\hat{\gamma}^{(j+1)} &= n \left[\sum_{i=1}^n \frac{1 + (1 - \hat{\gamma}^{(j)}) \exp(-\hat{\lambda}^{(j)} x_i)}{1 - (1 - \hat{\gamma}^{(j)}) \exp(-\hat{\lambda}^{(j)} x_i)} \right]^{-1} \\ \hat{\lambda}^{(j+1)} &= n \left[\sum_{i=1}^n \frac{x_i (1 + (1 - \hat{\gamma}^{(j)}) \exp(-\hat{\lambda}^{(j)} x_i))}{1 - (1 - \hat{\gamma}^{(j)}) \exp(-\hat{\lambda}^{(j)} x_i)} \right]^{-1}.\end{aligned}\quad (7)$$

Kitidamrongsuk (2010) shows in detail the computations of the expected Fisher information matrix $\mathbf{I}(\gamma, \lambda)$ of the EEG distribution, given by

$$\mathbf{I}(\gamma, \lambda) = \begin{bmatrix} \mathbf{I}_{11}(\gamma, \lambda) & \frac{n(\gamma - 1 - \gamma^2 \log(\gamma))}{3\lambda\gamma(\gamma - 1)^2} \\ \frac{n(\gamma - 1 - \gamma^2 \log(\gamma))}{3\lambda\gamma(\gamma - 1)^2} & \frac{n}{3\gamma^2} \end{bmatrix}, \quad (8)$$

with

$$\mathbf{I}_{11}(\gamma, \lambda) \begin{cases} \frac{n[3(1 - \gamma) - 2((1 - \gamma) - \gamma L_2(1 - \gamma))]}{3\lambda^2(1 - \gamma)} & , 0 < \gamma < 1 \\ \frac{n}{3\lambda^2(1 - \gamma)} \left[1 - \gamma \left(1 + \frac{\pi^2}{3} + \log^2(\gamma) - 2L_2\left(\frac{1}{\gamma}\right) \right) \right] & , \gamma > 1 \end{cases} \quad (9)$$

where

$$L_r(p) = \sum_{i=1}^{\infty} \frac{p^i}{i^r}, \quad (0 < p < 1)$$

is a polyarithmica function (Erdelyi et al., 1953). The maximum likelihood estimates for γ and λ are biased for small sample problems. In the case of large samples they become unbiased and asymptotically efficient. Such estimates are asymptotically Normal distributed with joint distribution given by

$$(\hat{\gamma}, \hat{\lambda}) \sim N_2[(\gamma, \lambda), I^{-1}(\gamma, \lambda)], \text{ when } n \rightarrow \infty. \quad (10)$$

Bayesian Analysis

In this section we consider the Bayesian estimation of the unknown parameters λ and γ .

First, a prior distribution which expresses little information on the parameters γ and λ can be obtained from uniform densities, which do not favor any particular value of λ and γ . In this case, the joint prior distribution for λ and γ is given by

$$\pi_U(\gamma, \lambda) \propto \text{constant}. \quad (11)$$

Another widely-used method to specify prior information is through the product of independent gamma distributions for each parameter λ and γ , since $\gamma > 0$ and $\lambda > 0$, that is, $\gamma \sim \text{Gamma}(\alpha_1, \beta_1)$ and $\lambda \sim \text{Gamma}(\alpha_2, \beta_2)$, where $\text{Gamma}(a, b)$ denotes a gamma distribution with mean a/b and variance a/b^2 ; and $\alpha_1, \alpha_2, \beta_1$ and β_2 are known hyperparameters. Thus, the joint prior distribution for λ and γ is given by

$$\pi_G(\gamma, \lambda) \propto \gamma^{\alpha_1-1} \lambda^{\alpha_2-1} \exp(-\beta_1 \gamma - \beta_2 \lambda). \quad (12)$$

Assume $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 0.01$, that is, a non-informative prior given by (12).

Another well-known existing non-informative prior, which represents a situation with little a priori information on the parameters was introduced by Jeffreys (1967), also known as the Jeffreys rule. The Jeffreys prior has been widely used due to the invariance property for one to one transformations of the parameters. The Jeffreys prior is defined as

$$\pi_J(\gamma, \lambda) \propto \sqrt{\det \mathbf{I}(\gamma, \lambda)}. \quad (13)$$

where $\mathbf{I}(\gamma, \lambda)$ is the Fisher information matrix defined in (8) and (9).

From the equation (13), we get

$$\begin{aligned}
 \pi_J(\gamma, \lambda) &\propto \sqrt{\frac{-2\left((\gamma-1)^3 L_2(1-\gamma) - (-\gamma + \gamma \log(\gamma) + 1)(\gamma^2 + \gamma^2 \log(\gamma) - 3\gamma + 2)\right)}{\gamma(\gamma-1)^4}} \times \\
 &\quad \times \frac{1}{\lambda}; \quad 0 < \gamma < 1, \\
 \pi_J(\gamma, \lambda) &\propto \sqrt{\frac{(\gamma-1)^3 \left(1 - \frac{\gamma}{3} \left(-6L_2\left(\frac{1}{\gamma}\right) + 3\log^2(\gamma) + \pi^2 + 3\right)\right) - (\gamma^2 \log(\gamma) - \gamma + 1)^2}{\gamma(\gamma-1)^4}} \times \\
 &\quad \times \frac{1}{\lambda}; \quad \gamma > 1.
 \end{aligned} \tag{14}$$

It is interesting to observe it was found in (14) independent priors for the parameters λ and γ , but this joint prior has a dependence structure.

Zellner (1977, 1984) proposed a non-informative prior based on the Shannon's entropy (1948). The idea is to maximize the information from the data in relation to the prior information on the parameters. This non-informative prior distribution known as "Maximal Data Information Prior" is obtained from the solution of the equation

$$\pi_Z(\gamma, \lambda) \propto \exp\left(\int_0^\infty \log(f(x|\gamma, \lambda)) f(x|\gamma, \lambda) dx\right). \tag{15}$$

For the EEG distribution given in (1) the resulting non-informative prior is given by

$$\pi_Z(\gamma, \lambda) \propto (\lambda\gamma) \exp\left(-\gamma\Psi(1-\gamma, 1, 1) + \frac{2\log(\gamma) + 2 - 2\gamma}{(\gamma-1)}\right), \tag{16}$$

where $\Psi(z, s, a)$ is defined in (5).

The proposed Zellner prior distribution (15) has limited invariance properties, where invariance is only verified under linear transformations of the vector (γ, λ) and not for all differentiable one by one transformations. Bernardo (1979) and Berger & Bernardo (1992) use the Kullback-Liebler distance between the posterior distribution $p(\theta|x)$ the prior distribution $\pi(\theta)$ to maximize the information from the data in relation to the known prior information for the parameters to find a non-informative prior. Additional information about the reference prior can be found in Bernardo (2005).

An important feature in this approach is the different treatment for interest and nuisance parameters when θ is a vector of parameters. In the presence of nuisance parameters, a typical case in this paper, one must establish an ordered parameterization with the parameter of interest singled out and then follow the procedure below. The algorithm of Berger and Bernardo (1992) to derive the reference prior can be described in four steps, as follows. We will present here the two-parameters case in details.

Let $\theta = (\theta_1, \theta_2)$ be the two parameters vector; θ_1 will be considered the parameter of interest and θ_2 is the nuisance parameter. The algorithm used to obtain the reference prior is given by

Step 1: Find the conditional reference prior $\pi_2(\theta_2 | \theta_1)$, assuming that θ_1 is given by

$$\pi_2(\theta_2 | \theta_1) = \sqrt{\mathbf{I}_{22}(\theta_1, \theta_2)}, \quad (17)$$

where $\mathbf{I}_{22}(\theta_1, \theta_2)$ is the term of order (2,2) of the information Fisher matrix.

Step 2: Normalize $\pi_2(\theta_2 | \theta_1)$. If $\pi_2(\theta_2 | \theta_1)$ is improper, choose a sequence of sets $\Omega_1 \subseteq \Omega_2 \subseteq \dots \rightarrow \Omega$, where $\pi_2(\theta_2 | \theta_1)$ is proper. Find

$$c_m(\theta_1) = \frac{1}{\int_{\Omega_m} \pi_2(\theta_2 | \theta_1) d\theta_2} \quad (18)$$

$$p_m(\theta_2 | \theta_1) = c_m(\theta_1) \pi_2(\theta_2 | \theta_1) 1_{\Omega_m}(\theta_2). \quad (19)$$

Step 3: Find the reference prior for θ_1 . The result is given by the solution of

$$\pi_m(\theta_1) \propto \exp \left\{ \frac{1}{2} \int_{\Omega_m} p_m(\theta_2 | \theta_1) \log \left\| \frac{\det \mathbf{I}(\theta_1, \theta_2)}{\mathbf{I}_{22}(\theta_1, \theta_2)} \right\| d\theta_2 \right\}. \quad (20)$$

Step 4: Find the prior distribution for (θ_1, θ_2) , when θ_2 is the nuisance parameter

$$\pi_2(\theta_1, \theta_2) = \lim_{m \rightarrow \infty} \left(\frac{c_m(\theta_1) \pi_m(\theta_1)}{c_m(\theta_1^*) \pi_m(\theta_1^*)} \right) \pi_2(\theta_2 | \theta_1), \quad (21)$$

where θ_1^* is any fixed point within the positive density for all π_m .

For the EEG distribution given in (1), the reference prior when λ is the parameter of interest is given by

$$\pi_\lambda(\lambda, \gamma) \propto \begin{cases} \sqrt{\frac{(1-\gamma) + 2\gamma L_2(1-\gamma)}{(1-\gamma)}} & 0 < \gamma < 1 \\ \sqrt{\frac{1}{(1-\gamma)} \left[1 - \gamma \left(1 + \frac{\pi^2}{3} + \log^2(\gamma) - 2L_2\left(\frac{1}{\gamma}\right) \right) \right]} & \gamma > 1 \end{cases} \quad (22)$$

The reference prior when γ is the parameter of interest is given by

$$\begin{aligned} \pi_\gamma(\lambda, \gamma) &\propto \sqrt{\frac{\gamma \left(-2 \left((\gamma-1)^3 L_2(1-\gamma) - (-\gamma + \gamma \log(\gamma) + 1)(\gamma^2 + \gamma^2 \log(\gamma) - 3\gamma + 2) \right) \right)}{(\gamma-1)^4}} \\ &\text{for } 0 < \gamma < 1 \text{ and} \\ \pi_\gamma(\lambda, \gamma) &\propto \sqrt{\frac{\gamma \left((\gamma-1)^3 \left(1 - \frac{\gamma}{3} \left(-6L_2\left(\frac{1}{\gamma}\right) + 3\log^2(\gamma) + \pi^2 + 3 \right) \right) - (\gamma^2 \log(\gamma) - \gamma + 1)^2 \right)}{(\gamma-1)^4}} \\ &\text{for } \gamma > 1 \end{aligned} \quad (23)$$

Finally, derive the prior distributions for the parameters the resulting joint posterior distributions for γ and λ is proportional to the product of the likelihood function (4) and the prior distributions $\pi(\gamma, \lambda)$ given in (11), (12), (14), (16), (22) and (23), that is

$$p(\gamma, \lambda | x) \propto \pi(\gamma, \lambda) (\lambda \gamma)^n \exp \left(-\lambda \sum_{i=1}^n x_i \right) \prod_{i=1}^n \left\{ \left(1 - (1-\gamma) e^{-\lambda x_i} \right)^{-2} \right\}. \quad (24)$$

By using any prior distribution proposed is not possible to derive the marginal posterior distributions in an analytical form for the parameters γ and λ . Thus, to obtain the posterior information on the parameters of interest as the point estimator and Bayes credibility intervals, we use MCMC algorithms to simulate samples of the values of γ and λ from the joint posterior distributions.

Results

Two applications of the theoretical results discussed in the previous sections are presented. The first involves a comparison of the estimation methods based on simulated data; the second shows an application of the EEG distribution to real data.

Analysis via numerical simulation

In this example, some simulations are performed via the Monte Carlo method. The goal is to study the effect of different non-informative prior distributions on the posterior summaries and also to compare these results with the obtained results using classical inference analysis. Posterior summaries of interest are evaluated using Monte Carlo Markov Chain (MCMC) methods. The influence of sample size on the accuracy of the obtained estimators is also examined. The following procedure was adopted:

1. Determine the values of γ and λ .
2. Specify the sample size n .
3. Generate values of a distribution $EEG(\gamma, \lambda)$ with size n .
4. Using the data obtained in Step 3, calculate the estimates for the parameters γ and λ using MCMC in the Bayesian approach and MLE in the classical approach.
5. Repeat the steps 3 and 4 N times.

Consider two set of the true values for the parameter (γ, λ) given by $(\gamma, \lambda) = (0.5, 2)$ and $(\gamma, \lambda) = (2, 4)$ representing decreasing and increasing hazard functions, respectively. The simulated data are generate from EEG distribution with the parameter values above for different sample sizes, as $n = 10, 25$ and 50 .

Tables 1 and 2 show the posterior mean and median, respectively, by considering the non-informative priors proposed in this paper for the parameters. The maximum likelihood estimates (MLE) are also available.

OBJECTIVE PRIORS FOR ESTIMATION OF GEOMETRIC DISTRIBUTION

Table 1. Posterior medians and MLE for $\lambda = 2$ and $\gamma = \frac{1}{2}$ for 1000 samples of sizes 10, 25 and 50.

$\lambda = 2$	Jeffreys	MDIP	Ref. λ	Ref. γ	Uniform	Gamma	MLE
n=10	2.98(1.62)	2.80(0.84)	2.96(0.87)	2.88(0.81)	3.61(1.54)	1.15(1.06)	3.44(2.61)
n=25	2.62(1.10)	3.00(1.05)	2.93(0.88)	2.92(0.83)	2.95(1.00)	2.01(1.06)	2.54(1.26)
n=50	2.17(0.76)	2.52(0.70)	2.58(0.67)	2.71(0.67)	2.56(0.72)	2.07(0.77)	2.26(0.80)

$\gamma = \frac{1}{2}$	Jeffreys	MDIP	Ref. λ	Ref. γ	Uniform	Gamma	MLE
n=10	1.01(0.43)	1.31(0.77)	1.12(0.17)	1.13(0.14)	1.50(0.77)	0.35(0.49)	1.51(2.02)
n=25	0.87(0.39)	0.92(0.42)	0.95(0.23)	0.99(0.21)	0.97(0.46)	0.56(0.41)	0.81(0.68)
n=50	0.63(0.32)	0.71(0.27)	0.81(0.27)	0.89(0.26)	0.76(0.32)	0.56(0.29)	0.64(0.36)

Table 2. Posterior medians and MLE for $\lambda = 4$ and $\gamma = 2$ for 1000 samples of sizes 10, 25 and 50.

$\lambda = 4$	Jeffreys	MDIP	Ref. λ	Ref. γ	Uniform	Gamma	MLE
n=10	4.85(1.21)	4.62(1.50)	5.31(1.55)	4.73(0.99)	5.08(1.25)	5.08(1.25)	4.86(2.03)
n=25	4.56(0.81)	4.02(0.77)	3.82(0.71)	3.90(0.73)	4.77(0.87)	3.50(0.84)	4.41(1.42)
n=50	3.76(0.57)	3.98(0.61)	3.51(0.41)	3.56(0.40)	3.40(0.64)	3.67(0.65)	4.22(0.94)

$\gamma = 2$	Jeffreys	MDIP	Ref. λ	Ref. γ	Uniform	Gamma	MLE
n=10	3.07(0.50)	1.81(0.40)	3.25(0.42)	3.33(0.40)	3.26(0.42)	3.26(0.42)	3.65(3.23)
n=25	2.70(0.49)	1.76(0.34)	1.62(0.25)	1.74(0.27)	2.94(0.53)	1.52(0.56)	2.87(2.06)
n=50	1.71(0.45)	1.90(0.41)	1.53(0.29)	1.60(0.32)	1.95(0.45)	1.70(0.52)	2.43(1.16)

From Tables 1 and 2, it is observed that when the hazard function is decreasing ($0 < \gamma < 1$) the prior distribution given by product of independent gamma distributions gives the best estimation for the parameters while for the increasing hazard function ($\gamma > 1$) the MDIP prior distribution provides the best one for all sample sizes considered.

A criterion for comparison of the prior distributions consists on checking the frequentist coverage probabilities of the posterior intervals. We therefore compare the frequency at which the true values of γ and λ are included in their 95% posterior intervals. This frequency should be close to 95% for large numbers of repeated experiments.

Table 3. Coverage probabilities for $\lambda = 2$ and $\gamma = \frac{1}{2}$ for 1000 samples of sizes 10, 25 and 50.

$\lambda = 2$	Jeffreys	MDIP	Ref. λ	Ref. γ	Uniform	Gamma	MLE
n=10	98.50%	99.10%	95.70%	96.70%	90.60%	96.50%	95.20%
n=25	95.20%	91.60%	89.70%	90.00%	93.40%	97.20%	95.00%
n=50	97.50%	94.30%	95.00%	91.30%	92.20%	96.60%	94.70%

$\gamma = \frac{1}{2}$	Jeffreys	MDIP	Ref. λ	Ref. γ	Uniform	Gamma	MLE
n=10	99.20%	98.00%	95.70%	94.80%	90.60%	97.10%	94.60%
n=25	95.50%	97.80%	97.90%	97.70%	95.00%	97.80%	95.50%
n=50	97.60%	97.90%	96.00%	95.50%	93.70%	97.70%	95.40%

Table 4. Coverage probabilities for $\lambda = 4$ and $\gamma = 2$ for 1000 samples of sizes 10, 25 and 50.

$\lambda = 4$	Jeffreys	MDIP	Ref. λ	Ref. γ	Uniform	Gamma	MLE
n=10	95.50%	96.60%	88.10%	98.00%	91.20%	91.20%	96.60%
n=25	94.60%	98.20%	96.80%	97.00%	94.00%	95.60%	94.30%
n=50	97.30%	98.40%	96.00%	97.60%	98.50%	96.60%	95.00%

$\gamma = 2$	Jeffreys	MDIP	Ref. λ	Ref. γ	Uniform	Gamma	MLE
n=10	97.10%	99.50%	93.80%	97.00%	96.60%	96.60%	92.50%
n=25	98.50%	99.10%	99.50%	99.90%	98.50%	97.10%	92.00%
n=50	98.20%	98.90%	97.00%	99.50%	98.80%	96.50%	93.90%

An example with literature data

Now consider a lifetime dataset related to an electrical insulator subjected to constant stress and strain, introduced by Lawless (1982). The dataset does not have censored values and represents the lifetime (in minutes) to failure: 0.96, 4.15, 0.19, 0.78, 8.01, 31.75, 7.35, 6.50, 8.27, 33.91, 32.52, 16.03, 4.85, 2.78, 4.67, 1.31, 12.06, 36.71 and 72.89.

Assume that the EEG distribution is appropriated to analyze this dataset, and then it will be compared with other lifetime distributions such as Weibull, Gamma, and Lognormal. As shown, the efficiency of the different non-informative prior distributions changes with the shapes of the hazard functions, therefore, to get good inferences on parameters of interest it is necessary to have some prior information on how the hazard function behaves for the Lawless data set. In this way, Barlow & Campo (1975) proposed a simple graphical technique that has been widely used to verify the behavior of the risk function called TTT plot (total time for testing). The graph is constructed with the plot of the consecutive quantities $[r/n, G(r/n)]$, where $G(r/n)$ is a function given by

$$G(r/n) = \frac{\sum_{i=1}^r t_i + (n-r)t_{(r)}}{\sum_{i=1}^n t_i},$$

where $r = 1, \dots, n$ and $t_{(i)}, i = 1, \dots, n$ are the order statistics in the sample.

Using the TTT curve in an empirical scale, one can determine the shape of the hazard function for the lifetime data. A diagonal line indicates that the data have a constant hazard function; if the curve is convex the risk is decreasing; if it is concave, there is an indication that the risk is increasing; if first is convex and after this is concave then there is an indication that there is a bathtub shape for the hazard function; if it is first concave and after this convex, there is an indication of inverse form of the bath for the hazard function. The Figure 3 shows how to verify the behavior of the hazard function.

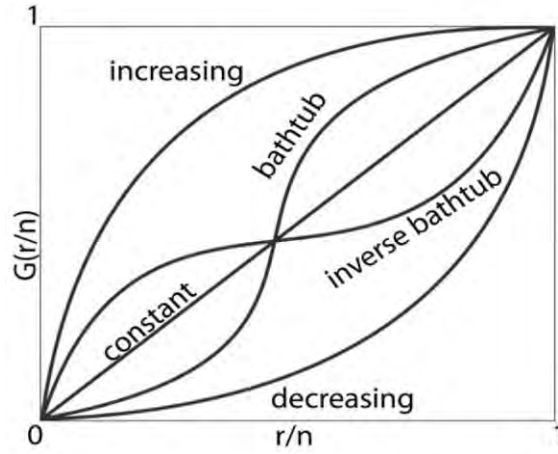


Figure 3. TTT plots for different distributions indicating the shape of the hazard function.

Some TTT transformations can be studied to solve other problems. Nair et al. (2008) show some of these transformations applied in survival analysis. Figure 4 shows the TTT plot for the Lawless data set.

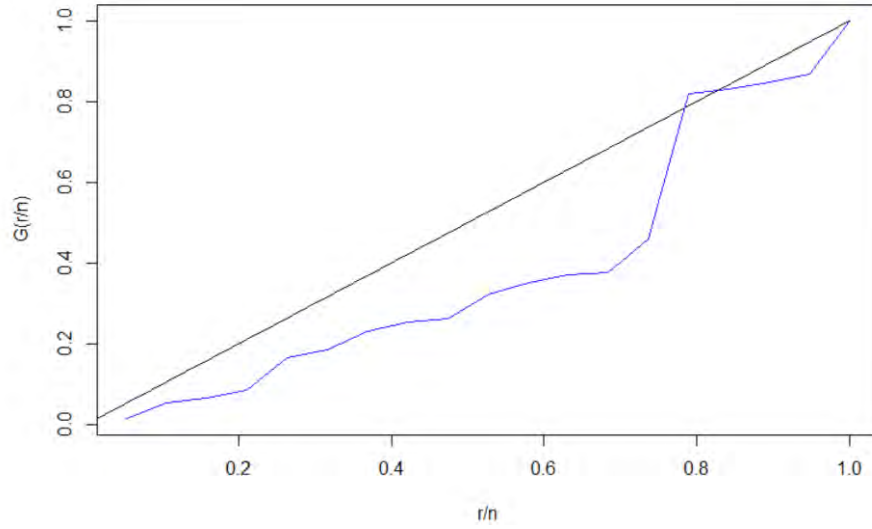


Figure 4. TTT plot for the dataset lifetime related to an electrical insulator subjected to constant stress and strain (Lawless data).

It is observed in Figure 4 that the TTT plot is convex; then it can be concluded that the risk is a decreasing function. When the hazard function is decreasing, it was observed from the results of section 4.1, that non-informative priors obtained through the product of independent gamma distributions is the best prior with little prior information about the parameters of interest. The joint posterior distribution of λ and γ (24) is obtained by replacing $\pi(\gamma, \lambda)$ by (12). It is necessary to use numerical methods to extract information from the marginal posterior distributions λ and γ . MCMC methods are used to simulate samples for the joint posterior distribution; that is, also for the marginal posterior distributions of interest.

It was generated 110,000 iterations with a “burn-in” of 10,000 values and jumps of size 10; so we get chains of the marginal posterior distributions for λ and γ of size 10,000 obtained using MCMC methods. To verify the convergence of the chains, we have used Geweke (1992) diagnosis, which indicated the convergence of the two chains. The convergence and autocorrelations is also observed in the trace-plots of the simulated series given in Figure 5.

To verify the performance of other lifetime distributions we also consider as non-informative prior, the product of independent gamma distributions $\gamma \sim \text{Gamma}(0.01, 0.01)$, $\lambda \sim \text{Gamma}(0.01, 0.01)$ assuming the following lifetime distributions: EGE, Weibull, Gamma and Lognormal distribution. The results are compiled in Tables 5 and 6, respectively.

OBJECTIVE PRIORS FOR ESTIMATION OF GEOMETRIC DISTRIBUTION

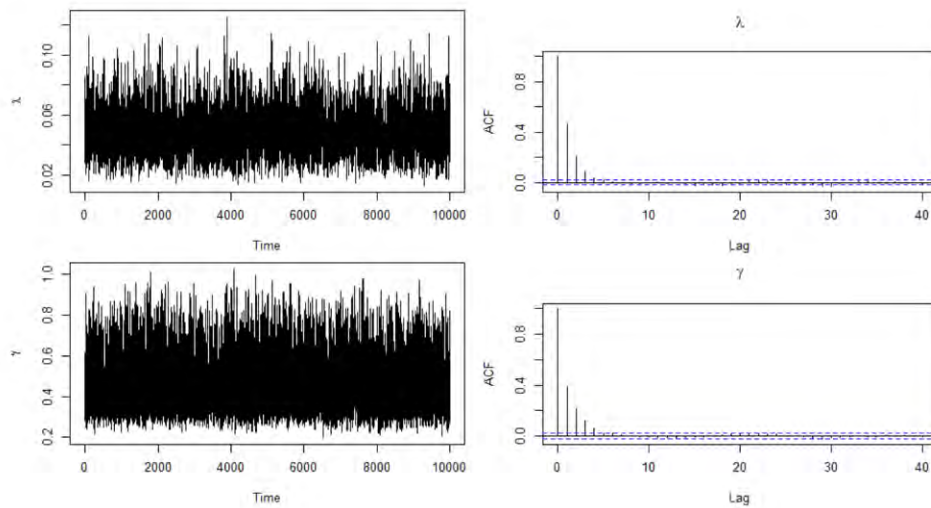


Figure 5. Trace-plots and autocorrelation graphs for the generated values of λ and γ .

Table 5. Posterior estimates (means) for γ and λ considering different probability distributions for the Lawless data set.

Parameters	EGE	Weibull	Gamma	Log-normal
γ	0.0482 (0.0161)	0.7629(0.1356)	0.6725(0.1870)	1.6880(0.3969)
λ	0.4513 (0.1649)	7.8250(4.2250)	0.0474(0.0183)	2.6410(1.0200)

Table 6. Obtained results for the DIC, BIC and AIC criteria for the different probability distributions for the Lawless data set.

Criteria	EGE	Weibull	Gamma	Log-normal
DIC	138.96	140.70	141.30	141.30
BIC	140.55	148.06	143.13	151.62
AIC	142.44	149.95	141.24	149.73

Based on any of the criteria used by the table it can be concluded that EGE was the best fit to the offered data.

Conclusion

The use of extended exponential geometric (EEG) distributions showed a good flexibility of fit for lifetime data applications and could be an alternative

distribution to other usual distributions in Survival analysis. The great number of existing non-informative prior distributions can cause difficulties in the choice of an adequate prior with little information a priori, mainly when these prior distributions do not produce similar posterior summaries. In this way, the development of a general theory for the construction of non-informative prior distributions is an important topic to be investigated by researchers in the Bayesian inference.

The results showed the effects of different non-informative prior distributions related to the changes in the risk function using extended exponential geometric (EEG) distributions. Therefore, we recommend the product of gamma distributions Gamma (0.01, 0.01) when the hazard function is decreasing and the non-informative MDIP prior distribution when the hazard function is increasing. With these choices of prior we surely get better inferences for the parameters.

References

- Achcar, J. A, Moala, F. A, & Boleta, J. (March, 2010). *Generalized Exponential Distribution: A Bayesian approach using MCMC methods*. Poster presented at the 10th Bayesian Statistics Brazilian Meeting, Rio de Janeiro, Brazil.
- Adamidis, K, Dimitrakopoulou, T, & Loukas, S. (2005). On an Extension of the Exponential Geometric Distribution. *Statistics and Probability Letters*, 73, 259-269.
- Adamidis, K, & Loukas, S. (1998). A Lifetime Distribution with Decreasing Failure Rate. *Statistics and Probability Letters*, 39, 35-42.
- Barlow, R. E., & Campo, R. A. (1975). Total Time on Test processes and applications to failure data analysis. In R. E. Barlow, J. B. Fussel, & N. D. Singpurwalla (Eds.), *Reliability and Fault Tree Analysis: Theoretical and Applied Aspects of System Reliability and Safety Assessment* (451-481). Philadelphia, PA: Society for Industrial and Applied Mathematics.
- Bayes, T. R. (1958). Essay towards solving a problem in the doctrine of changes. Reprinted in *Biometrika*, 45, 243-315, 1763.
- Berger, J. O, & Bernardo, J. M. (1992). *On the Development of the Reference Prior Method*. Fourth Valencia International Meeting on Bayesian Statistics, Spain.

- Bernardo, J. M. (1979). Reference Posterior Distributions for Bayesian Inference. *Journal Royal Statistical Society*, 41(2), 113-147.
- Bernardo, J. M. (2005). Reference analysis. In D.K. Dey and C. R. Rao, (Eds.), *Handbook of Statistics*, 25 (17-90). Amsterdam: Elsevier.
- Box, G. E. P., & Tiao, G. C. (1973). *Bayesian inference in statistical analysis*. Reading, MA: Addison-Wesley.
- Chib, S., & Greenberg, E. (1995). Understanding The Metropolis-Hasting Algorithm. *The American Statistician*, 49(4), 327-335.
- Degroot, M. H., & Schervish, M. J. (2002). *Probability and Statistics*. Reading, MA: Addison-Wesley.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm (with discussion). *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), 1-38.
- Erdelyi, A., Maguns, W., Oberhettinger, F., & Tricomi, F.G. (1953). *Higher Transcendental Functions*. New York: McGraw-Hill.
- Gelfand, A. E., & Smith, F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85, 398-409.
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In J. M. Bernado, J. O. Berger, A. P. Dawid, & A. F. M. Smith (Eds.), *Bayesian Statistics 4*. Oxford, UK: Clarendon Press.
- Gupta, R. D., & Kundu, D. (1999). Generalized exponential distributions. *Australian and New Zealand Journal of Statistics*, 41, 173-188.
- Gupta, R. D., & Kundu, D. (2001). Generalized exponential distributions: different methods of estimation. *Journal of Statistical Computation and Simulation*, 69, 315-338.
- Gupta, R. D., & Kundu, D. (2007). Generalized exponential distribution: existing results and some recent developments. *Journal of Statistical Planning and Inference*, 137(11), 3537-3547. doi: 10.1016/j.jspi.2007.03.030.
- Gupta, R. D., & Kundu, D. (2008). Generalized exponential distribution; Bayesian Inference. *Computational Statistics and Data Analysis*, 52(4), 1873-1883.
- Hamada, M., Wilson, A. G., Reese, C. S., & Martz, H. F. (2008). *Bayesian Reliability*. New York: Springer.
- Jeffreys, H. (1967). *Theory of probability* (3rd ed.). London: Oxford University Press.

- Kitidamrongsuk, P. (2010). *Discriminating Between the Extended Exponential Geometric Distribution and the Gamma Distribution*. (Unpublished Doctoral Dissertation). Assumption University of Thailand.
- Laplace, P. (1774). Mémoire sur la probabilité des causes par les événements. *Mem. Acad. R. Sci. Présentés par Divers Savans*, v. 6, p. 621-656 (translated in *Statistical Science*, 1, 359-378).
- Lawless, J. F. (1982). *Statistical models and methods for lifetime data*. New York: John Wiley and Sons.
- Marshall, A.W., & Olkin, I. (1997). A New Method for Adding a Parameter to a Family of Distributions with Application to the Exponential and Weibull Families. *Biometrika*, 84, 641-652.
- Moala, F. A. (2010). Bayesian analysis for the Weibull parameters by using noninformative prior distributions. *Advances and Applications in Statistics*, 14(2), 117-143.
- Nair, N. U., Sankaran, P. G., & Kumar, B. V., (2008). Total time on test transforms of order n and its implications in reliability analysis. *Journal of Applied Probability*, 45(4), 1126–1139.
- Raqab, M. Z. (2002). Inferences for generalized exponential distribution based on record statistics. *Journal of Statistical Planning and Inference*, 104, 339-350.
- Raqab, M. Z., & Ahsanullah, M. (2001). Estimation of the location and scale parameters of generalized exponential distribution based on order statistics. *Journal of Statistical Computation and Simulation*, 69, 109-124.
- Sarhan, A. M. (2007). Analysis of incomplete, censored data in competing risks models with generalized exponential distributions. *IEEE Transactions on Reliability*, 56(1), 132-138.
- Shannon, C. E. (1948). A Mathematical theory of communication. *Bell System Technical Journal*, 27, 623-659.
- Zellner, A. (1977). Maximal Data Information Prior Distributions. In A. Aykac & C. Brumat (Eds.), *New Methods in the applications of Bayesian Methods*. Amsterdam: North-Holland Publishing Co.
- Zellner, A. (1984). *Maximal Data Information Prior Distributions: Basic Issues in Econometrics*. Chicago, IL: University of Chicago Press.
- Zheng, G. (2002). Fisher information matrix in type-II censored data from exponentiated exponential family. *Biometrical Journal*, 44, 353-357.

Discrete Generalized Burr-Type XII Distribution

B. A. Para

University of Kashmir
Srinagar, India

T. R. Jan

University of Kashmir
Srinagar, India

A discrete analogue of generalized Burr-type XII distribution is introduced using a general approach of discretizing a continuous distribution. It may be worth exploring the possibility of developing a discrete version of the six parameter generalized Burr-type XII distribution for use in modeling a discrete data. This distribution is suggested as a suitable reliability model to fit a range of discrete lifetime data, as it is shown that hazard rate function can attain monotonic increasing (decreasing) shape for certain values of parameters. The equivalence of discrete generalized Burr-type XII (DGBD-XII) and continuous generalized Burr-type XII (GBD-XII) distributions has been established. The increasing failure rate property in the discrete setup has been ensured. Various theorems relating this new model to other probability distributions have also been proved.

Keywords: Discrete generalized Burr-type XII distribution, discrete lifetime models, reliability, failure rate

Introduction

In reliability theory a number of continuous life models is now available in the subject to portray the survival behavior of a component or a system. Many continuous life distributions have been studied in details (see for example [Kapur & Lamberson, 1997](#); [Lawless, 1982](#); [Sinha, 1986](#)). However, it is sometimes impossible or inconvenient in life testing experiments to measure the life length of a device on a continuous scale. For example the lifetime of an on/off switching device is a discrete random variable, or life length of a device receiving a number of shocks it sustain before it fails is also a discrete random variable.

Recently, the special roles of discrete distributions have received recognition from survival analysts. Many continuous distributions have been discretized. For example, the Geometric and Negative binomial distributions are the discrete

B. A. Para is a Research Scholar in the Department of Statistics. Email him at parabilal@gmail.com. T. R. Jan is an Assistant Professor in the Post-Graduate Department of Statistics. Email him at drtrjan@gmail.com.

versions of Exponential and Gamma distributions. Nakagawa (1975) discretized the Weibull distribution. The discrete versions of the normal and Rayleigh distributions were also proposed by Roy (2003, 2004). Discrete analogues of Maxwell, two parameter Burr-type XII and Pareto distributions were also proposed by Krishna and Punder (2007, 2009). Recently the inverse Weibull distribution was also discretized by Jazi, Lai and Alamatsaz (2010). This article addresses the problem of discretization of generalized Burr-type XII (GBD-XII) distribution, because there is a need to find more plausible discrete lifetime distributions to fit to various life time data.

The Discrete Generalized Burr XII Distribution:

Roy (1993) pointed out that the univariate geometric distribution can be viewed as a discrete concentration of a corresponding exponential distribution in the following manner

$$p[X = x] = s(x) - s(x+1) \text{ when } x = 0, 1, 2, \dots$$

where X is discrete random variable following geometric distribution with probability mass functions as

$$p(x) = \theta^x (1 - \theta), x = 0, 1, 2, \dots$$

where $s(x)$ represents the survival function of an exponential distribution of the form $s(x) = \exp(-\lambda x)$, clearly $\theta = \exp(-\lambda)$, $0 < \theta < 1$. Thus, one to one correspondences between the geometric distribution and the exponential distribution can be established, the survival functions being of the same form.

The general approach of discretizing a continuous variable is to introduce a greatest integer function of X i.e., $[X]$ (the greatest integer less than or equal to X till it reaches the integer), in order to introduce grouping on a time axis. If the underlying continuous failure time X has the survival function $s(x) = p(X > x)$ and times are grouped into unit intervals, so that the discrete observed variable is $dX = [X]$.

DISCRETE GENERALIZED BURR-TYPE XII DISTRIBUTION

The probability mass function of dX can be written as

$$\begin{aligned} p(x) &= p(dX = x) = p(x \leq X < x+1) \\ &= \varnothing(x+1) - \varnothing(x) \\ &= s(x) - s(x+1), \quad x = 0, 1, 2, \dots \end{aligned}$$

$\varnothing(x)$ being the cumulative distribution function of random variable X .

In reliability theory, many classification properties and measures are directly related to the functional form of the survival function. The increasing failure rate (IFR), decreasing failure rate (DFR), Increasing failure rate average (IFRA), decreasing failure rate average (DFRA), new better than used (NBU), new worse than used (NWU), new better (worse) than used in expectation NBUE (NWUE) and increasing (decreasing) mean residual lifetime IMRL (DMRL) etc. are examples of such class properties as may be seen from Barlow and Proschan (1975). If discretization of a continuous life distribution can retain the same functional form of the survival function then many reliability measures and class properties will remain unchanged. In this sense, the discrete concentration concept is considered herein as a simple approach that can generate a discrete life distribution model. Thus, given any continuous life variable with survival function $s(x)$ a discrete lifetime variable X with probability mass function $p(x)$ is defined by

$$p(x) = s(x) - s(x+1), x = 0, 1, 2, \dots$$

Using this concept for the purpose of discretizing generalized Burr-type XII distribution, suppose that Y_1 and Y_2 are independently distributed continuous random variables. If Y_1 has an exponential density function $f(y_1, \theta) = \theta e^{-\theta y_1}, y_1 > 0; \theta > 0$ and Y_2 has a gamma distribution with pdf

$$f(y_2; \alpha, k) = \frac{\alpha^k}{\Gamma(k)} e^{-\alpha y_2} y_2^{k-1}, \alpha > 0; k > 0; y_2 > 0, \text{ then the random variable}$$

$X = c \sqrt{\frac{Y_1}{Y_2}}$ has a six parameter generalized Burr-type XII distribution with parameters $(\mu = 0, \sigma = 1, \alpha, \theta, c, k)$ with a density function

$$f(x) = \frac{k\theta\alpha c \alpha^k x^{c-1}}{(\alpha + \theta x^c)^{k+1}}; \quad x > 0, \theta > 0; c > 0; \alpha > 0; k > 0 \quad (1)$$

The pdf plot for DGB-XII variate X for different values of parameters is shown in Figure 1. It is evident that the distribution of the random variable X is right skewed.

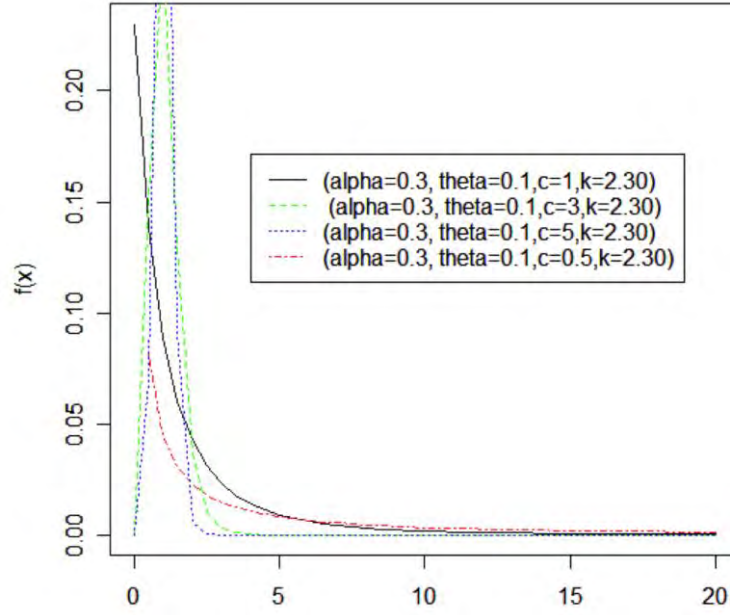


Figure 1. PDF plot for GBD-XII (α, θ, c, k)

Introducing location parameter μ and scale parameter σ in (1) results in

$$f(x_1) = \frac{k\theta c \alpha^k \left(\frac{x_1 - \mu}{\sigma} \right)^{c-1}}{\sigma \left[\alpha + \theta \left(\frac{x_1 - \mu}{\sigma} \right)^c \right]^{k+1}}, \quad x > 0, \alpha > 0, c > 0, \theta > 0, \mu > 0, \sigma > 0, k > 0$$

Reliability measures of GBD-XII random variable X

Various reliability measures of GBD-XII random variable X are given by

Survival Function:

$$\begin{aligned} s(x) &= p(X \geq x) = 1 - \int_0^x f(x) dx \\ &= 1 - \int_0^x \frac{x^k k \theta c x^{c-1}}{(x + \theta x^c)^{k+1}} dx \\ &= \alpha^k (\alpha + \theta x^c)^{-k} \end{aligned}$$

r^{th} Moment:

$$E(x^r) = k \left(\frac{\alpha}{\theta} \right)^{r/c} B\left(\frac{r}{c} + 1, k - r/c \right), x > 0; k > r/c; \alpha > 0; \theta > 0; r > 0; c > 0$$

$$\text{where } B(a, b) = \int_0^\infty \frac{x^{a-1}}{(1+x)^{a+b}} dx$$

Failure Rate:

$$r(x) = \frac{f(x)}{s(x)} = \frac{k \theta c x^{c-1}}{(\alpha + \theta x^c)}, x > 0; k > 0; \alpha > 0; \theta > 0; c > 0$$

Second Rate of Failure:

$$\begin{aligned} SRF(x) &= \log \left(\frac{S(x)}{S(x+1)} \right) \\ &= -k \log \left[\frac{\alpha + \theta x^c}{\alpha + \theta (x+1)^c} \right] \quad x > 0; k > 0; \alpha > 0; \theta > 0; c > 0 \end{aligned}$$

A discrete generalized Burr XII variable, dX can be viewed as the discrete concentration of the continuous generalized Burr-type XII variate X distribution, where the corresponding probability mass function of dX can be written as

$$\begin{aligned}
 p(dX = x) &= p(x) = s(x) - S(x+1) \\
 &= \alpha^k (\alpha + \theta x^c)^{-k} - \alpha^k (\alpha + \theta (x+1)^c)^{-k} \\
 &= \alpha^{-\log \beta} \left[\beta^{\log(\alpha + \theta x^c)} - \beta^{\log(\alpha + \theta (x+1)^c)} \right]
 \end{aligned}$$

The pmf of DGBD-XII takes the form

$$p(x) = \begin{cases} 1 - \alpha^{\log \beta} \beta^{(\alpha + \theta)} & x = 0 \\ \alpha^{-\log \beta} \left(\beta^{\log(\alpha + \theta x^c)} - \beta^{\log(\alpha + \theta (x+1)^c)} \right) & x = 1, 2, 3, \dots \end{cases} \quad (2)$$

$$\begin{aligned}
 \text{where } \beta &= e^{-k}; & 0 < \beta < 1 \\
 & & k > 0; \alpha > 0; c > 0; \beta > 0
 \end{aligned}$$

Figures 2 through 5 give the pmf plot of (2) for $(\alpha = 0.3; \theta = 0.1; c = 4; \beta = 0.1)$, $(\alpha = 0.3; \theta = 0.1; c = 1; \beta = 0.1)$, $(\alpha = 0.3; \theta = 0.01; c = 0.5; \beta = 0.1)$, and $(\alpha = 1; \theta = 1; c = 2; \beta = 0.5)$, respectively.

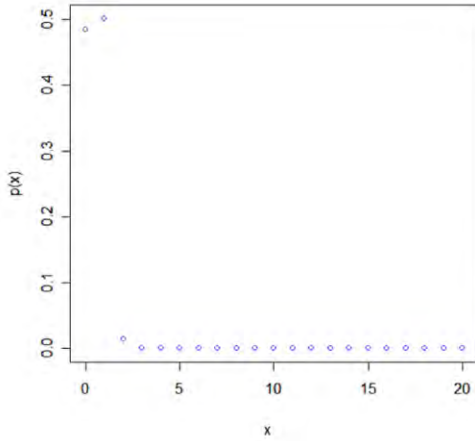


Figure 2. PMF plot for DGBD-XII
 $(\alpha = 0.3; \theta = 0.1; c = 4; \beta = 0.1)$

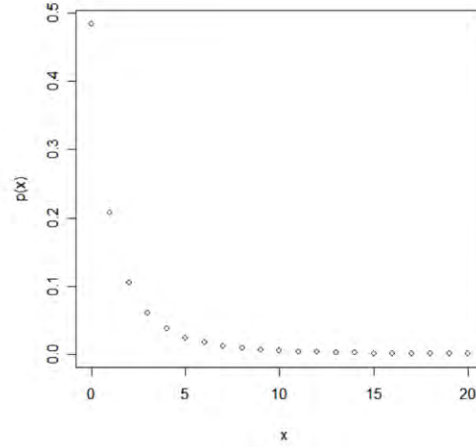


Figure 3. PMF plot for DGBD-XII
 $(\alpha = 0.3; \theta = 0.1; c = 1; \beta = 0.1)$

DISCRETE GENERALIZED BURR-TYPE XII DISTRIBUTION

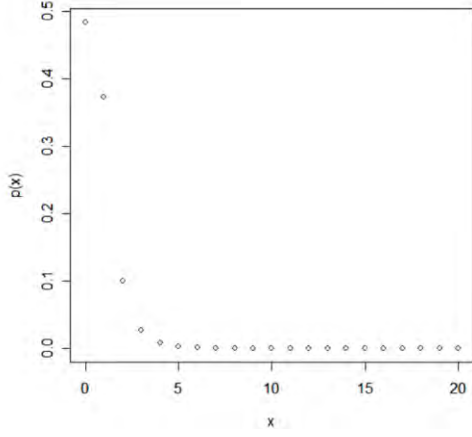


Figure 4. PMF plot for DGBD-XII
($\alpha = 0.3$; $\theta = 0.01$; $c = 0.5$; $\beta = 0.1$)

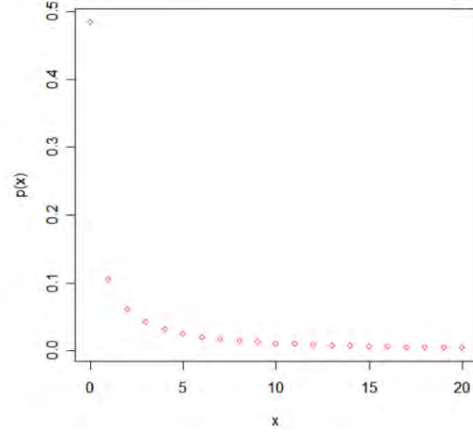


Figure 5. PMF plot for DGBD-XII
($\alpha = 1$; $\theta = 1$; $c = 2$; $\beta = 0.5$)

The pmf at $x = 0$ is independent of the shape parameter c . It is monotonic decreasing if

$$c \leq \frac{\log\left(\frac{e^{\varnothing(\beta, \alpha, \theta)} - \alpha}{\theta}\right)}{\log 2} \quad \text{where } \varnothing(\beta, \alpha, \theta) = \frac{\log(2\beta^{\log(\alpha+\theta)} - \alpha^{\log \beta})}{\log \beta}$$

where $\beta = e^{-k}$; $0 < \beta < 1$; $k > 0$; $\alpha > 0$; $c > 0$; $\theta > 0$

otherwise it is no longer monotonic decreasing but is unimodal, having a mode at $x = 1$ i.e., it takes a jump at $x = 1$ and then decreases for all $x \geq 1$. For $\alpha = \theta = c = 1$ pmf of discrete generalized Burr-type XII distribution coincides with discrete pareto distribution and for $\alpha = \theta = 1$ DGBD-XII reduces to discrete Burr-type XII distribution.

To introduce location parameter μ and scale parameter σ then the discretized version of $f(x_1)$ is given as

$$\begin{aligned}
 P(dX_1 = x_1) &= p(x_1) = p(x_1 \leq X_1 < x_1 + 1) \\
 &= P(x_1 \leq \sigma x + \mu < x_1 + 1) \\
 &= P\left(\frac{x_1 - \mu}{\sigma} \leq X < \frac{x_1 + 1 - \mu}{\sigma}\right) \\
 &= \Phi\left(\frac{x_1 + 1 - \mu}{\sigma}\right) - \Phi\left(\frac{x_1 - \mu}{\sigma}\right)
 \end{aligned}$$

where $\Phi(x_1)$ represents cumulative distribution function of random variable X .

The survival function of discrete generalized Burr-type XII random variable dX is given by

$$\begin{aligned}
 s(x) &= p(dX \geq x) = 1 - p(dX < x) \\
 &= 1 - [p(dX = 0) + p(dX = 1) + \dots + p(dX = x - 1)] \\
 &= a^k \beta^{\log(x + \theta x^c)} = \alpha^{-\log \beta} \beta^{\log(x + \theta x^c)} \quad x = 0, 1, 2, \dots \\
 &\quad a > 0; c > 0; k > 0; \theta > 0
 \end{aligned}$$

Thus, survival function of discrete generalized DGBD-XII is same as continuous GBD-XII for integer points of x .

The failure rate of discrete generalized Burr-type XII random variable is given by

$$\begin{aligned}
 r(x) &= \frac{p(x)}{s(x)} = 1 - \beta^{\log \left[\frac{\alpha + \theta(x+1)^c}{(\alpha + \theta x^c)} \right]} \\
 &\quad x = 0, 1, 2, \dots; a > 0; c > 0; k > 0; \theta > 0
 \end{aligned} \tag{3}$$

i.e., the conditional probability that failure occurs at a time x given that the system has not failed by $x - 1$.

Note that $r(0) = r(1)$ gives $c = \frac{\log(\frac{\theta}{\alpha} + 2)}{\log 2} = a$ (for example). If $\alpha = \theta = 1$, the

hazard function of discrete generalized Burr-type XII distribution coincides with two parameter discrete Burr-type XII distribution. Note that $r(x)$ is decreasing in x if $0 < c < a$ and for $c = a$; $r(0) = r(1)$ and for $c > a$, $r(0) > r(1)$ i.e., monotonic increasing, and $r(x) < r(x - 1) \forall x > 2$ i.e., $r(x)$ decreases for all $x > 1$, uniformly in β .

DISCRETE GENERALIZED BURR-TYPE XII DISTRIBUTION

For different values of α and θ , c can take different values as:

- a) $\theta = .5$; $\alpha = 100$ gives $c = 1.003$
- b) $\theta = .99$; $\alpha = .5$ gives $c = 1.993$
- c) $\theta = 9999$; $\alpha = .9$ gives $c = 13.43$
- d) $\theta = 1$; $\alpha = 1$ gives $c = 1.585$

Taking part a) it could be seen that for

$\theta = .5$; $\alpha = 100$ gives $c = 1.003$ (for example as above)

$r(0) > r(1)$, for $0 < c < a$ (for example when $c = 1$)

and for $c = a$; $r(0) = r(1)$

and for $c > a$, $r(0) < r(1)$ implies a monotonic increasing

and $r(x) < r(x - 1) \forall x \geq 2$

i.e., $r(x)$ takes a jump at $r = 1$ and decreases for all $x \geq 1$, uniformly in β .

Similarly, for all other cases where c can take different values for different values of α and θ , $r(x)$ will show its monotonicity accordingly as in the above case.

For discrete generalized Burr-type XII distribution i.e., DGBD-XII(α, θ, c, β)

$$\begin{aligned} SRF(x) &= \log \left(\frac{s(x)}{s(x+1)} \right) = \log \left(\frac{\alpha^{-\log \beta} \beta^{\log(\alpha + \theta x^c)}}{\alpha^{-\log \beta} \beta^{\log(\alpha + \theta (x+1)^c)}} \right) \\ &= \log \beta \log \left(\frac{\alpha + \theta x^c}{\alpha + \theta (x+1)^c} \right), \quad \alpha > 0; c > 0; k > 0; \theta > 0 \end{aligned}$$

and to see whether $SRF(x)$ shows the same monotonicity as $r(x)$

$$SRF(x) = \log \beta \log \left(\frac{\alpha + \theta x^c}{\alpha + \theta (x+1)^c} \right)$$

For $SRF(0) = SRF(1)$

$$c = \log_2 \left(\frac{\theta}{\alpha} + 2 \right) = \frac{\log \left(\frac{\theta}{\alpha} + 2 \right)}{\log 2} = \alpha \text{ (for example)}$$

Using the same procedure as in (3) it is clear that $SRF(x)$ shows the same monotonicity as that of $r(x)$.

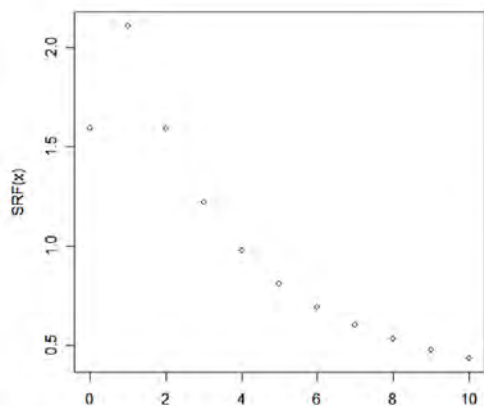


Figure 6. second rate of failure
plot of DGBD-XII
($\alpha = 1$; $\theta = 1$; $c = 2$; $\beta = 0.1$)

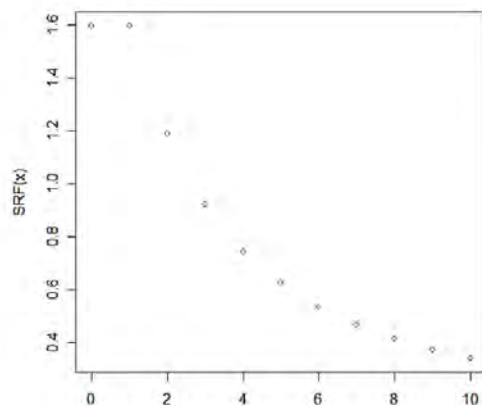


Figure 7. second rate of failure
plot of DGBD-XII
($\alpha = 1$; $\theta = 1$; $c = 1.585$; $\beta = 0.1$)

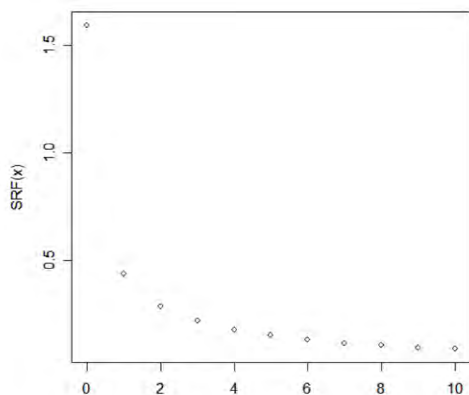


Figure 8. second rate of failure
plot of DGBD-XII
($\alpha = 1$; $\theta = 1$; $c = 0.5$; $\beta = 0.1$)

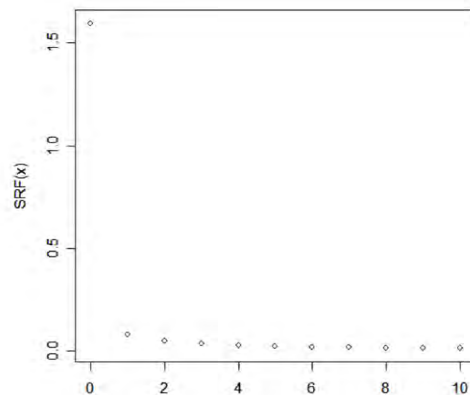


Figure 9. second rate of failure
plot of DGBD-XII
($\alpha = 1$; $\theta = 1$; $c = 0.1$; $\beta = 0.1$)

Figures 6 through 9 illustrate the alternative hazard rate plot for DGBD-XII for different values of parameters. Note that $SRF(x)$ is decreasing in x if $0 < c < a$

DISCRETE GENERALIZED BURR-TYPE XII DISTRIBUTION

and for $c = a$; $SRF(0) = SRF(1)$ and for $c > a$, $SRF(0) > SRF(1)$ i.e., monotonic increasing, and $SRF(x) < SRF(x - 1) \forall x > 2$ i.e., $SRF(x)$ decreases for all $x > 1$.

The r^{th} moment of discrete generalized Burr-type XII distribution, i.e. DGBD $(\alpha, \theta, c, \beta)$, is as follows

$$\begin{aligned} E(X^r) &= \sum_{x=0}^{\infty} x^r p(x) = \sum_{x=0}^{\infty} x^r \left[\beta^{\log(\alpha+\theta x^c)} - \beta^{\log(\alpha+\theta c+x)^c} \right] \alpha^{-\log \beta} \\ &= \sum_{x=1}^{\infty} \left[x^r - (x-1)^r \right] S(x) \end{aligned} \quad (4)$$

$$E(X^r) \leq r \sum_{x=1}^{\infty} x^{r-1} s(x) \leq \frac{r \alpha^{-\log \beta}}{\theta^k} \sum_{x=1}^{\infty} \frac{1}{x^{ck-r+1}}$$

Now, $E(x^r) \leq r \sum_{x=1}^{\infty} \frac{r \alpha^{-\log \beta}}{\theta^k} \sum_{x=1}^{\infty} \frac{1}{x^{ck-r+1}}$, $E(x^r)$ will take a finite value if $ck > r$,

and from (4), $E(x) = \sum_{x=1}^{\infty} s(x) = \sum_{x=1}^{\infty} \alpha^{-\log \beta} \beta^{\log(\alpha+\theta x^c)}$ and

$$\begin{aligned} E(x^2) &= \sum_{x=1}^{\infty} \left[x^2 - (x-1)^2 \right] s(x) \\ &= 2 \sum_{x=1}^{\infty} x s(x) - E(x) \\ E(x^2) &= 2 \sum_{x=1}^{\infty} \alpha^{-\log \beta} \beta^{\log(\alpha+\theta x^c)} + E(x) \end{aligned}$$

which are infinite series and cannot be written as closed form. The parameter β of DGBD-XII $(\alpha, \theta, c, \beta)$ and the parameter k of GBD-XII (α, θ, c, k) , are matched via $\beta = e^{-k}$. It is therefore observed from the survival functions of DGBD-XII and GBD-XII distributions

$$\sum_{x=1}^{\infty} \left\{ \alpha^{-\log \beta} \beta^{\log(\alpha+\theta x^c)} \right\} < \int_0^{\infty} \alpha^k (\alpha + \theta x^c)^{-k} dx < \sum_{x=0}^{\infty} \left\{ \alpha^{-\log \beta} \beta^{-\log \beta} \beta^{\log(\alpha+\theta x^c)} \right\}$$

In other words, $\mu_d - 1 < \mu_c < \mu_d$ where μ_c and μ_d are the means of the continuous and discrete generalized Burr XII distributions, respectively.

References

- Barlow, R. E., & Proschan, F. (1975). *The statistical theory of reliability and life testing*. New York: Holt, Rinehart & Winston.
- Jazi, M. A., Lai, C. D., & Alamatsaz, M. H. (2010). A discrete inverse Weibull distribution and estimation of its parameters. *Statistical Methodology*, 7, 121–132.
- Kapur, K. C., & Lamberson, L. R. (1997). *Reliability in engineering design*. New York: John Wiley & Sons.
- Krishna, H., & Pundir, P. S. (2007). Discrete Maxwell Distribution. *Interstat*. Retrieved from <http://interstat.statjournals.net/YEAR/2007/abstracts/0711003.php>
- Krishna, H., & Pundir, P. S. (2009). Discrete Burr and discrete Pareto distributions. *Statistical Methodology*, 6, 177–188.
- Lawless, J. F. (1982). *Statistical models and methods for lifetime data*. New York: John Wiley & Sons.
- Nakagawa, T., & Osaki, S. (1975). The discrete Weibull distribution. *IEEE Transactions on Reliability*, R-24(5), 300–301.
- Olapade, A. K. (2008). On a six parameter generalized Burr XII distribution. Retrieved from <http://arxiv.org/abs/0806.1579v1>
- Roy, D. (1993). Reliability measures in the discrete bivariate set up and related characterization results for a bivariate geometric distribution. *Journal of Multivariate Analysis*, 46, 362–373.
- Roy, D. (2003). The discrete normal distribution. *Communications in Statistics – Theory and Methods*, 32(10), 1871–1883.
- Roy, D. (2004). Discrete Rayleigh distribution. *IEEE Transactions on Reliability*, 53(2), 255–260.
- Sinha, S. K. (1986). *Reliability and life testing*. New Delhi: Wiley Eastern Ltd.
- Srinath, L. S. (1985). *Reliability engineering* (4th ed.). New Delhi: East-West Press.
- Xie, M., Gaudoin, M., & Bracquemond, C. (2002). Redefining failure rate function for discrete distributions. *International Journal of Reliability, Quality and Safety Engineering*, 9(3), 275–285.

Appendix: Some theorems related to discrete generalized Burr XII distribution

Lemma 1

If X is a continuous random variable with increasing (decreasing) failure rate IFR (DFR) distribution, then $dX = [X]$ has a discrete increasing (decreasing) failure rate dIFR (dDFR).

Proof: (See Roy and Dasgupta, 2001)

Lemma 2

If X is a non-negative continuous random variable and Y is a non-negative integer valued discrete random variable, then

$$[X] \geq Y \Leftrightarrow X \geq Y$$

Proof: Note that

$$([X] \geq Y) \subseteq (X \geq Y) \subseteq ([X] \geq [Y]) = ([X] \geq Y)$$

where the last equality holds because Y is integer valued. Therefore

$$(X \geq Y) = ([X] \geq Y)$$

Theorem 1

If $X \sim \text{DGBD-XII} (\mu = 0; \sigma = 1; \alpha, \theta, c, \beta)$ then

$$Y = \left\lceil \log \left(\frac{\alpha + \theta x^c}{\alpha} \right) \right\rceil \sim \text{Geo}(\beta) \quad \text{where } \beta = e^{-k}; \alpha > 0; c > 0; k > 0; \theta > 0$$

Proof:

$$\begin{aligned}
 P[Y \geq y] &= P\left[\left[\log\left(\frac{\alpha + \theta x^c}{\alpha}\right)\right] \geq y\right] \\
 &= P\left[\log\left(\frac{\alpha + \theta x^c}{\alpha}\right) \geq y\right] && \text{by Lemma 2} \\
 P[Y \geq y] &= P\left[X \geq \left(\frac{e^{y - \log \alpha^{-1}} - \alpha}{\theta}\right)^{1/c}\right] \\
 &= \alpha^{-\log \beta} \beta^{\left[\log\left\{\alpha + \theta \left(\frac{e^{y - \log \alpha^{-1}} - \alpha}{\theta}\right)^{1/c}\right\}^c\right]} \\
 &= \alpha^{-\log \beta} \beta^{y - \log \alpha^{-1}} \\
 &= \beta^y \sim \text{Geo}(\beta) && y = 0, 1, 2, \dots
 \end{aligned}$$

As β^y is the survival function of geometric random variable.

Theorem 2

If $X \sim \text{GBD-XII} (\mu = 0; \sigma = 1; \alpha, \theta, c, k)$ then $Y = [X] \sim \text{DGBD-XII} (\mu = 0; \sigma = 1; \alpha, \theta, c, \beta)$; where $\beta = e^{-k}$; $\alpha > 0$; $c > 0$; $k > 0$; $\theta > 0$

Proof:

$$\begin{aligned}
 P[Y \geq y] &= P[[X] \geq y] = P[X \geq y] && \text{by Lemma 2} \\
 &= \alpha^k (\alpha + \theta y^c)^{-k} \\
 &= \alpha^{-\log \beta} \beta^{\log(\alpha + \theta y^c)} && y = 0, 1, 2, \dots \\
 &&& \beta = e^{-k}; 0 < \beta < 1
 \end{aligned}$$

Thus, $Y = [X] \sim \text{DGBD-XII} (\alpha, \theta, c, \beta)$

Theorem 3

If $X \sim \exp(k)$, the exponential distribution with failure rate k . Then

$$Y = \left[\left(\frac{e^{X+\log \alpha} - \alpha}{\theta} \right)^{1/c} \right] \sim \text{DGBD-XII} (\alpha, \theta, c, \beta), \text{ where } \beta = e^{-k}; \alpha > 0, c > 0, k > 0, \\ \theta > 0.$$

Proof:

$$\begin{aligned} P[Y \geq y] &= P \left[\left[\left(\frac{e^{X+\log \alpha} - \alpha}{\theta} \right)^{1/c} \right] \geq y \right] \\ &= P \left[\left(\frac{e^{X+\log \alpha} - \alpha}{\theta} \right)^{1/c} \geq y \right] && \text{by Lemma 2} \\ &= P \left[X \geq \log(\theta y^c + \alpha) - \log \alpha \right] \\ &= e^{-k[\log(\theta y^c + \alpha) - \log \alpha]} \\ &= \alpha^k e^{-k \log(\alpha + \theta y^c)} \\ P[Y \geq y] &= \alpha^{-\log \beta} \beta^{\log(\alpha + \theta y^c)} && y = 0, 1, 2, \dots \end{aligned}$$

which is the survival function of DGBD-XII $(\alpha, \theta, c, \beta)$. Thus, $Y \sim \text{DGBD-XII} (\alpha, \theta, c, \beta)$.

Theorem 4

Let X be a random variable following continuous generalized Burr XII distribution with $E(X^r) < \infty \forall r = 1, 2, \dots$

Then $E(Y^r) < \infty$ where $Y = [X] \sim \text{DGB} (\alpha, \theta, c, k)$

Proof: Proof is straightforward, because $0 \leq [X] \leq X$, so clearly if $E(X^r) < \infty \forall r = 1, 2, \dots$, then $E([X]^2) < \infty$.

Bayesian Estimation of the Parameters of Two-Component Mixture of Rayleigh Distribution under Doubly Censoring

Tabassum Naz Sindhu
Quaid-i-Azam University
Islamabad, Pakistan

Navid Feroze
Allama Iqbal Open University
Islamabad, Pakistan

Muhammad Aslam
Quaid-i-Azam University
Islamabad, Pakistan

Recently, the Bayesian analysis of the two-component mixture of lifetime models under singly type I censored samples was discussed. The Bayes estimation of the parameters of mixture of two Rayleigh distributions (MTRD) is developed under doubly censoring. Different informative priors, under squared error loss function and k-loss function, have been assumed for the posterior estimation. The performance of different estimators has been compared in terms of posterior risks by analyzing the simulated and real life data sets.

Keywords: Inverse transformation method, mixture model, doubly censoring, loss functions, Bayes estimator

Introduction

In survival analysis, data are subject to censoring. The most common type of censoring is right censoring, in which the survival time is larger than the observed right censoring time. In some cases, however, data are subject to left, as well as, right, censoring. When left censoring occurs, the only information available to an analyst is that the survival time is less than or equal to the observed left censoring time. A more complex censoring scheme is found when both initial and final times are interval-censored. This situation is referred as double censoring, or the data with both right and left censored observations are known as doubly censored data.

Analysis of doubly censored data for simple (single) distribution has been studied by many authors. Fernandez (2000) investigated maximum likelihood prediction based on type II doubly censored exponential data. Fernandez (2006) has discussed Bayesian estimation based on trimmed samples from Pareto populations. Khan et al. (2010) studied predictive inference from a two-parameter Rayleigh life

Mr. Feroze is in the Department of Statistics. Email at navidferoz@gmail.com. Mr. Sindhu and Mr. Aslam are in the Department of Mathematics and Statistics.

model given a doubly censored sample. Kim and Song (2010) have discussed Bayesian estimation of the parameters of the generalized exponential distribution from doubly censored samples. Khan et al. (2011) studied sensitivity analysis of predictive modeling for responses from the three-parameter Weibull model with a follow-up doubly censored sample of cancer patients. Pak et al. (2013) has proposed the estimation of Rayleigh scale parameter under doubly type-II censoring from imprecise data.

In statistics, a mixture distribution is signified as a convex fusion of other probability distributions. It can be used to model a statistical population with subpopulations, where constituent of mixture probability densities are the densities of the subpopulations. Mixture distribution may appropriately be used for certain data set where the subsets of the whole data set possess different properties that can best be modeled separately. They can be more mathematically manageable, because the individual mixture components are dealt with more ease than the overall mixture density. The families of mixture distributions have a wider range of applications in different fields such as fisheries, agriculture, botany, economics, medicine, psychology, electrophoresis, finance, communication theory, geology and zoology.

Soliman (2006) derived estimators for the finite mixture of Rayleigh model based on progressively censored data. Sultan, et al. (2007) described the properties and estimation of mixture of two inverse Weibull distributions. Sultan, et al. (2007) have discussed some properties of the mixture of two inverse Weibull distributions. Saleem and Aslam (2008) presented a comparison of the Maximum Likelihood (ML) estimates with the Bayes estimates assuming the Uniform and the Jeffreys priors for the parameters of the Rayleigh mixture. Kundu and Howalder (2010) considered the Bayesian inference and prediction of the inverse Weibull distribution for type-II censored data. Saleem et al. (2010) considered the Bayesian analysis of the mixture of Power function distribution using the complete and the censored sample. Shi and Yan (2010) studied the case of the two parameter exponential distribution under type I censoring to get empirical Bayes estimates. Eluebaly and Bouguila (2011) have presented a Bayesian approach to analyze finite generalized Gaussian mixture models which incorporate several standard mixtures, widely used in signal and image processing applications, such as Laplace and Gaussian. Sultan and Al-Moisheer (2012) developed approximate Bayes estimation of the parameters and reliability function of mixture of two inverse Weibull distributions under Type-2 censoring.

The Proposed Mixture Model and the Likelihood Function

The probability density function (pdf) of the Rayleigh distribution with rate parameter λ_i is

$$f_i(x_{ij}) = 2x_{ij}\lambda_i^2 \exp(-x_{ij}^2\lambda_i^2), \quad 0 < x_{ij} < \infty, \quad \lambda_i^2 > 0, \quad i = 1, 2, \text{ and } j = 1, 2, \dots, n_i \quad (1)$$

The cumulative distribution function (CDF) of the distribution is

$$F_i(x_{ij}) = 1 - \exp(-\lambda_i^2 x_{ij}^2), \quad 0 < x_{ij} < \infty, \quad \lambda_i^2 > 0, \quad i = 1, 2, \text{ and } j = 1, 2, \dots, n_i \quad (2)$$

A density function for mixture of two components densities with mixing weights $(p_1, 1-p_1)$ is

$$f(x) = p_1 f_1(x) + (1-p_1) f_2(x), \quad 0 < p_1 < 1. \quad (3)$$

The cumulative distribution function for the mixture model is:

$$F(x) = p_1 F_1(x) + (1-p_1) F_2(x) \quad (4)$$

Consider a random sample of size ‘ n ’ from Rayleigh distribution, and let x_r, x_{r+1}, \dots, x_s be the ordered observations that can only be observed. The remaining ‘ $r-1$ ’ smallest observations and the ‘ $n-s$ ’ largest observations have been assumed to be censored. Now based on causes of failure, the failed items are assumed to come either from subpopulation 1 or from subpopulation 2; so the $x_{1r_1}, \dots, x_{1s_1}$ and $x_{2r_2}, \dots, x_{2s_2}$ failed items come from first and second subpopulations respectively. The rest of the observations which are less than x_r and greater than x_s have been assumed to be censored from each component. Where $x_s = \max(x_{1s_1}, x_{2s_2})$ and $x_r = \min(x_{1r_1}, x_{2r_2})$. Therefore, $m_1 = s_1 - r_1 + 1$ and $m_2 = s_2 - r_2 + 1$ number of failed items can be observed from first and second subpopulations respectively. The remaining $n - (s - r + 2)$ items are assumed to be censored observations, and $s - r + 2$ are the uncensored items. Where $r = r_1 + r_2$,

$s = s_1 + s_2$ and $m = m_1 + m_2$. Then the likelihood function for the Type II doubly censored sample $\mathbf{x} = \left\{ (x_{1r_1}, \dots, x_{1s_1}), (x_{2r_2}, \dots, x_{2s_2}) \right\}$, assuming the causes of the failure of the left censored items are identified, can be written as:

$$L(\lambda_1, \lambda_2, p_1 | \mathbf{x}) \propto p_1^{s_1} (1 - p_1)^{s_2} \left\{ F_1(x_{(r_1)}, \lambda_1) \right\}^{r_1-1} \left\{ F(x_{(r_2)}, \lambda_2) \right\}^{r_2-1} \left\{ 1 - F(x_s, \lambda_1, \lambda_2) \right\}^{n-s} \left\{ \prod_{i=r_1}^{s_1} f_1(x_{1(i)}, \lambda_1) \right\} \left\{ \prod_{i=r_2}^{s_2} f_2(x_{2(i)}, \lambda_2) \right\} \quad (5)$$

$$L(\lambda_1, \lambda_2, p_1 | \mathbf{x}) \propto \sum_{k_1=0}^{r_1-1} \sum_{k_2=0}^{r_2-1} \sum_{k_3=0}^{n-s} (-1)^{k_1+k_2} \binom{r_1-1}{k_1} \binom{r_2-1}{k_2} \binom{n-s}{k_3} p_1^{n-s-k_3+s_1} (1-p_1)^{s_2+k_3} \times \lambda_1^{2m_1} \lambda_2^{2m_2} \exp \left\{ -\lambda_1^2 \left(\Omega(x_{1j}) \right) \right\} \exp \left\{ -\lambda_2^2 \left(\Omega(x_{2j}) \right) \right\} \quad (6)$$

where $W(x_{1j}) = \sum_{i=r_1}^{s_1} x_{1(i)}^2 + (n-s-k_3)x_{(s)}^2 + kx_{(r_1)}^2$, $W(x_{2j}) = \sum_{i=r_2}^{s_2} x_{2(i)}^2 + k_3x_{(s)}^2 + kx_{(r_2)}^2$,

$m_1 = s_1 - r_1 + 1$, and $m_2 = s_2 - r_2 + 1$

Bayes Estimation

For the Bayesian estimation, let us assume that the parameters λ_i and p_i $i=1, 2$ are independent random variables, and then we consider the following priors for different parameters:

Bayesian Estimation using Nakagami Prior

The prior for the rate parameters λ_i for $i = 1, 2$, is assumed to be the Nakagami distribution, with the hyper-parameters a_i and b_i , given by

$$f_{\lambda_i}(\lambda_i) = \frac{2a_i^{a_i}}{\Gamma(a_i)b_i^{a_i}} \lambda_i^{2a_i-1} \exp\left(\frac{-\lambda_i^2 a_i}{b_i}\right), \quad a_i, b_i > 0 \quad (7)$$

The prior for p_1 is assumed to be the beta distribution, whose density is given by

$$f_p(p_1) = \frac{\Gamma(c_1 + d_1)}{\Gamma(c_1)\Gamma(d_1)} p_1^{c_1-1} (1-p_1)^{d_1-1}, \quad c_1, d_1 > 0 \quad (8)$$

From equation (7)-(8), propose the following joint prior density of the vector $\Theta = (\lambda_1, \lambda_2, p_1)$

$$g(\Theta) \propto \lambda_i^{2a_i-1} \exp\left(\frac{-\lambda_i^2 a_i}{b_i}\right) p_1^{c_1-1} (1-p_1)^{d_1-1}, \quad (9)$$

$$0 < p_1 < 1, a_i > 0, b_i > 0, c_1 > 0, d_1 > 0$$

By multiplying Equation (9) with Equation (6), the joint posterior density for the vector Θ given the data becomes

$$\pi(\Theta | x) \propto \sum_{k_1=0}^{r_1-1} \sum_{k_2=0}^{r_2-1} \sum_{k_3=0}^{n-s} \prod_{i=1}^2 (-1)^{k_1+k_2} \binom{r_1-1}{k_1} \binom{r_2-1}{k_2} \binom{n-s}{k_3} p_1^{n-s-k_3+s_1+c_1-1} \quad (10)$$

$$\times (1-p_1)^{s_2+k_3+d_1-1} \lambda_i^{2(a_i+m_i)-1} \exp\left\{-\lambda_i^2 \left(\frac{a_i}{b_i} + \Omega(x_{ij})\right)\right\}$$

Marginal distributions of λ_i and p_1 $i=1,2$ can be obtained by integrating the nuisance parameters.

Bayesian Estimation using Chi Prior

The prior for the rate parameters λ_i for $i=1,2$, is assumed to be the chi distribution, with the hyperparameter e_i , given by

$$f_{\lambda_i}(\lambda_i) = \frac{2^{\frac{1-e_i}{2}}}{\Gamma\left(\frac{e_i}{2}\right)} \lambda_i^{e_i-1} \exp\left(-\frac{\lambda_i^2}{2}\right), \quad e_i > 0 \quad (11)$$

The prior for p_1 is assumed to be the beta distribution, whose density is given by

$$f_p(p_1) = \frac{\Gamma(c_2 + d_2)}{\Gamma(c_2)\Gamma(d_2)} p_1^{c_2-1} (1-p_1)^{d_2-1}, \quad c_2, d_2 > 0 \quad (12)$$

From equation (11)-(12), we propose the following joint prior density of the vector $\Theta = (\lambda_1, \lambda_2, p_1)$

$$g(\Theta) \propto \lambda_i^{e_i-1} \exp\left(-\frac{\lambda_i^2}{2}\right) p_1^{c_2-1} (1-p_1)^{d_2-1}, \quad 0 < p_1 < 1, e_i > 0, c_2 > 0, d_2 > 0 \quad (13)$$

By multiplying Equation (13) with Equation (6), the joint posterior density for the vector Θ given the data becomes

$$\begin{aligned} \pi(\Theta | x) \propto & \sum_{k_1=0}^{r_1-1} \sum_{k_2=0}^{r_2-1} \sum_{k_3=0}^{n-s} \prod_{i=1}^2 (-1)^{k_1+k_2} \binom{r_1-1}{k_1} \binom{r_2-1}{k_2} \binom{n-s}{k_3} p_1^{n-s-k_3+s_1+c_2-1} \\ & \times (1-p_1)^{s_2+k_3+d_2-1} \lambda_i^{2m_i+e_i-1} \exp\left\{-\lambda_i^2 \left(\frac{1}{2} + \Omega(x_{ij})\right)\right\} \end{aligned} \quad (14)$$

Bayesian Estimation using Rayleigh Prior

The prior for the rate parameters λ_i for $i=1, 2$, is assumed to be the Rayleigh distribution, with the hyperparameter v_i given by

$$f_{\lambda_i}(\lambda_i) = \frac{\lambda_i}{v_i^2} \exp\left(-\frac{\lambda_i^2}{2v_i^2}\right), \quad v_i > 0 \quad (15)$$

The prior for p_1 is assumed to be the beta distribution, whose density is given by

$$f_p(p_1) = \frac{\Gamma(c_3 + d_3)}{\Gamma(c_3)\Gamma(d_3)} p_1^{c_3-1} (1-p_1)^{d_3-1}, \quad c_3, d_3 > 0 \quad (16)$$

From equation (15)-(16), propose the following joint prior density of the vector $\Theta = (\lambda_1, \lambda_2, p_1)$:

$$g(\Theta) \propto \lambda_i \exp\left(\frac{-\lambda_i^2}{2v_i^2}\right) p_1^{c_3-1} (1-p_1)^{d_3-1}, \quad 0 < p_1 < 1, \quad v_i, c_3 > 0, \quad d_3 > 0 \quad (17)$$

By multiplying Equation (17) with Equation (6), the joint posterior density for the vector Θ given the data becomes

$$\begin{aligned} \pi(\Theta | x) \propto & \sum_{k_1=0}^{r_1-1} \sum_{k_2=0}^{r_2-1} \sum_{k_3=0}^{n-s} \prod_{i=1}^2 (-1)^{k_1+k_2} \binom{r_1-1}{k_1} \binom{r_2-1}{k_2} \binom{n-s}{k_3} p_1^{n-s-k_3+s_1+c_3-1} \\ & \times (1-p_1)^{s_2+k_3+d_3-1} \lambda_i^{2(m_i)+1} \exp\left\{-\lambda_i^2 \left(\frac{1}{2v_i^2} + \Omega(x_{ij})\right)\right\} \end{aligned} \quad (18)$$

Marginal distributions of λ_i and p_1 $i = 1, 2$ can be obtained by integrating the nuisance parameters.

Bayes Estimation of the Vector of Parameters Θ

The Bayesian point estimation is connected to a loss function in general, signifying the loss induced when the estimate $\hat{\theta}$ differ from true parameter θ . Because there is no specific rule that helps us to identify the appropriate loss function to be used, squared error loss is used in this article as it serve as standard loss. It is well known that under the squared error loss function, the Bayes estimator of a function of the parameters is the posterior mean of the function and risk is the posterior variance. It is defined as $l(\hat{\theta}, \theta) = (\theta - \hat{\theta})^2$.

It was originally used in estimation problems when the unbiased estimator of θ was being considered. Another reason for its popularity is due to its relationship to least squares theory. The use of SELF makes the calculations simpler.

BAYESIAN ESTIMATION OF TWO-COMPONENT MIXTURE

The K-Loss function (KLF), defined as: $l(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2 / \hat{\theta}\theta$, was proposed by Wasan (1970). It is well fitted for a measure of inaccuracy for an estimator of a scale parameter of a distribution defined on $R^+ = (0, \infty)$. Under K-Loss function the Bayes estimates and posterior risks are defined as $\hat{\theta} = \sqrt{E(\theta | \mathbf{x}) / E(\theta^{-1} | \mathbf{x})}$, and $\rho(\hat{\theta}) = 2\{E(\theta | \mathbf{x})E(\theta^{-1} | \mathbf{x}) - 1\}$ respectively.

The respective marginal distribution of each parameter has been used to derive the Bayes estimators and posterior risks for λ_1, λ_2 and p_1 under the squared error loss function (SELF) and K- loss functions (KLF). The Bayes estimators and posterior risks of λ_1, λ_2 and p_1 under squared error loss function (SELF) assuming Nakagami prior are given as:

The Bayes estimators of λ_1, λ_2 and p_1 are:

$$\hat{\lambda}_{1(SELF)} = N^{-1} \sum_{k_1=0}^{r_1-1} \sum_{k_2=0}^{r_2-1} \sum_{k_3=0}^{n-s} (-1)^{k_1+k_2} \binom{r_1-1}{k_1} \binom{r_2-1}{k_2} \binom{n-s}{k_3} \frac{B(A_1, A_2) \Gamma(a_1 + m_1 + 1/2) \Gamma(a_2 + m_2)}{2\{a_1 / b_1 + \Omega(x_{1j})\}^{(a_1+m_1+1/2)} 2\{a_2 / b_2 + \Omega(x_{2j})\}^{(a_2+m_2)}}$$

$$\hat{\lambda}_{2(SELF)} = N^{-1} \sum_{k_1=0}^{r_1-1} \sum_{k_2=0}^{r_2-1} \sum_{k_3=0}^{n-s} (-1)^{k_1+k_2} \binom{r_1-1}{k_1} \binom{r_2-1}{k_2} \binom{n-s}{k_3} \frac{B(A_1, A_2) \Gamma(a_1 + m_1) \Gamma(a_2 + m_2 + 1/2)}{2\{a_1 / b_1 + \Omega(x_{1j})\}^{(a_1+m_1)} 2\{a_2 / b_2 + \Omega(x_{2j})\}^{(a_2+m_2+1/2)}}$$

$$\hat{p}_{1(SELF)} = N^{-1} \sum_{k_1=0}^{r_1-1} \sum_{k_2=0}^{r_2-1} \sum_{k_3=0}^{n-s} (-1)^{k_1+k_2} \binom{r_1-1}{k_1} \binom{r_2-1}{k_2} \binom{n-s}{k_3} \frac{B(A_1 + 1, A_2) \Gamma(a_1 + m_1) \Gamma(a_2 + m_2)}{2\{a_1 / b_1 + \Omega(x_{1j})\}^{(a_1+m_1)} 2\{a_2 / b_2 + \Omega(x_{2j})\}^{(a_2+m_2)}}$$

The Posterior risks of λ_1, λ_2 and p_1 are:

$$\rho\left(\hat{\lambda}_{1(SELF)}\right)=N^{-1}\sum_{k_1=0}^{r_1-1}\sum_{k_2=0}^{r_2-1}\sum_{k_3=0}^{n-s}(-1)^{k_1+k_2}\binom{r_1-1}{k_1}\binom{r_2-1}{k_2}\binom{n-s}{k_3}\frac{B\left(A_1,A_2\right)\Gamma\left(a_1+m_1+1\right)\Gamma\left(a_2+m_2\right)}{2\left\{a_1/b_1+\Omega\left(x_{1j}\right)\right\}^{\left(a_1+m_1+1\right)}2\left\{a_2/b_2+\Omega\left(x_{2j}\right)\right\}^{\left(a_2+m_2\right)}}-\left(\hat{\lambda}_{1(SELF)}\right)^2$$

$$\rho\left(\hat{\lambda}_{2(SELF)}\right)=N^{-1}\sum_{k_1=0}^{r_1-1}\sum_{k_2=0}^{r_2-1}\sum_{k_3=0}^{n-s}(-1)^{k_1+k_2}\binom{r_1-1}{k_1}\binom{r_2-1}{k_2}\binom{n-s}{k_3}\frac{B\left(A_1,A_2\right)\Gamma\left(a_1+m_1\right)\Gamma\left(a_2+m_2+1\right)}{2\left\{a_1/b_1+\Omega\left(x_{1j}\right)\right\}^{\left(a_1+m_1\right)}2\left\{a_2/b_2+\Omega\left(x_{2j}\right)\right\}^{\left(a_2+m_2+1\right)}}-\left(\hat{\lambda}_{2(SELF)}\right)^2$$

$$\rho\left(\hat{p}_{1(SELF)}\right)=N^{-1}\sum_{k_1=0}^{r_1-1}\sum_{k_2=0}^{r_2-1}\sum_{k_3=0}^{n-s}(-1)^{k_1+k_2}\binom{r_1-1}{k_1}\binom{r_2-1}{k_2}\binom{n-s}{k_3}\frac{B\left(A_1+2,A_2\right)\Gamma\left(a_1+m_1\right)\Gamma\left(a_2+m_2\right)}{2\left\{a_1/b_1+\Omega\left(x_{1j}\right)\right\}^{\left(a_1+m_1\right)}2\left\{a_2/b_2+\Omega\left(x_{2j}\right)\right\}^{\left(a_2+m_2\right)}}-\left(\hat{p}_{1(SELF)}\right)^2$$

The Bayes estimators of λ_1, λ_2 and p_1 under KLF are:

$$\hat{\lambda}_{1(KLF)} = \frac{\sum_{k_1=0}^{r_1-1}\sum_{k_2=0}^{r_2-1}\sum_{k_3=0}^{n-s}(-1)^{k_1+k_2}\binom{r_1-1}{k_1}\binom{r_2-1}{k_2}\binom{n-s}{k_3}\frac{B\left(A_1,A_2\right)\Gamma\left(a_1+m_1+1/2\right)\Gamma\left(a_2+m_2\right)}{2\left\{a_1/b_1+\Omega\left(x_{1j}\right)\right\}^{\left(a_1+m_1+1/2\right)}2\left\{a_2/b_2+\Omega\left(x_{2j}\right)\right\}^{\left(a_2+m_2\right)}}}{\sum_{k_1=0}^{r_1-1}\sum_{k_2=0}^{r_2-1}\sum_{k_3=0}^{n-s}(-1)^{k_1+k_2}\binom{r_1-1}{k_1}\binom{r_2-1}{k_2}\binom{n-s}{k_3}\frac{B\left(A_1,A_2\right)\Gamma\left(a_1+m_1-1/2\right)\Gamma\left(a_2+m_2\right)}{2\left\{a_1/b_1+\Omega\left(x_{1j}\right)\right\}^{\left(a_1+m_1-1/2\right)}2\left\{a_2/b_2+\Omega\left(x_{2j}\right)\right\}^{\left(a_2+m_2\right)}}}\right)^{\frac{1}{2}}$$

$$\hat{\lambda}_{2(KLF)} = \frac{\sum_{k_1=0}^{r_1-1}\sum_{k_2=0}^{r_2-1}\sum_{k_3=0}^{n-s}(-1)^{k_1+k_2}\binom{r_1-1}{k_1}\binom{r_2-1}{k_2}\binom{n-s}{k_3}\frac{B\left(A_1,A_2\right)\Gamma\left(a_1+m_1\right)\Gamma\left(a_2+m_2+1/2\right)}{2\left\{a_1/b_1+\Omega\left(x_{1j}\right)\right\}^{\left(a_1+m_1\right)}2\left\{a_2/b_2+\Omega\left(x_{2j}\right)\right\}^{\left(a_2+m_2+1/2\right)}}}{\sum_{k_1=0}^{r_1-1}\sum_{k_2=0}^{r_2-1}\sum_{k_3=0}^{n-s}(-1)^{k_1+k_2}\binom{r_1-1}{k_1}\binom{r_2-1}{k_2}\binom{n-s}{k_3}\frac{B\left(A_1,A_2\right)\Gamma\left(a_1+m_1\right)\Gamma\left(a_2+m_2-1/2\right)}{2\left\{a_1/b_1+\Omega\left(x_{1j}\right)\right\}^{\left(a_1+m_1\right)}2\left\{a_2/b_2+\Omega\left(x_{2j}\right)\right\}^{\left(a_2+m_2-1/2\right)}}}\right)^{\frac{1}{2}}$$

$$\hat{p}_{\mathbf{1}(KLF)} = \left(\frac{\sum_{k_1=0}^{r_1-1} \sum_{k_2=0}^{r_2-1} \sum_{k_3=0}^{n-s} (-1)^{k_1+k_2} \binom{r_1-1}{k_1} \binom{r_2-1}{k_2} \binom{n-s}{k_3} \frac{B(A_1+1, A_2) \Gamma(a_1+m_1) \Gamma(a_2+m_2)}{2 \{a_1/b_1 + \Omega(x_{1j})\}^{(a_1+m_1)} 2 \{a_2/b_2 + \Omega(x_{2j})\}^{(a_2+m_2)}}}{\sum_{k_1=0}^{r_1-1} \sum_{k_2=0}^{r_2-1} \sum_{k_3=0}^{n-s} (-1)^{k_1+k_2} \binom{r_1-1}{k_1} \binom{r_2-1}{k_2} \binom{n-s}{k_3} \frac{B(A_1-1, A_2) \Gamma(a_1+m_1) \Gamma(a_2+m_2-1/2)}{2 \{a_1/b_1 + \Omega(x_{1j})\}^{(a_1+m_1)} 2 \{a_2/b_2 + \Omega(x_{2j})\}^{(a_2+m_2)}}} \right)^{\frac{1}{2}}$$

The Posterior risks of λ_1, λ_2 and p_1 under KLF are:

$$\rho(\hat{\lambda}_{(KLF)}) = 2 \left[N^{-2} \sum_{k_1=0}^{r_1-1} \sum_{k_2=0}^{r_2-1} \sum_{k_3=0}^{n-s} (-1)^{k_1+k_2} \binom{r_1-1}{k_1} \binom{r_2-1}{k_2} \binom{n-s}{k_3} \frac{B(A_1, A_2) \Gamma(a_1 + m_1 + 1/2) \Gamma(a_2 + m_2)}{2\{a_1/b_1 + \Omega(x_{1j})\}^{(a_1+m_1+1/2)} 2\{a_2/b_2 + \Omega(x_{2j})\}^{(a_2+m_2)}} \right. \\ \left. \times \sum_{k_1=0}^{r_1-1} \sum_{k_2=0}^{r_2-1} \sum_{k_3=0}^{n-s} (-1)^{k_1+k_2} \binom{r_1-1}{k_1} \binom{r_2-1}{k_2} \binom{n-s}{k_3} \frac{B(A_1, A_2) \Gamma(a_1 + m_1 - 1/2) \Gamma(a_2 + m_2)}{2\{a_1/b_1 + \Omega(x_{1j})\}^{(a_1+m_1-1/2)} 2\{a_2/b_2 + \Omega(x_{2j})\}^{(a_2+m_2)}} - 1 \right]$$

$$\rho(\hat{\lambda}_{2(KLF)}) = 2 \left[N^{-2} \sum_{k_1=0}^{r_1-1} \sum_{k_2=0}^{r_2-1} \sum_{k_3=0}^{n-s} (-1)^{k_1+k_2} \binom{r_1-1}{k_1} \binom{r_2-1}{k_2} \binom{n-s}{k_3} \frac{B(A_1, A_2) \Gamma(a_1 + m_1) \Gamma(a_2 + m_2 + 1/2)}{2\{a_1/b_1 + \Omega(x_{1j})\}^{(a_1+m_1)} 2\{a_2/b_2 + \Omega(x_{2j})\}^{(a_2+m_2+1/2)}} \right. \\ \left. \times \sum_{k_1=0}^{r_1-1} \sum_{k_2=0}^{r_2-1} \sum_{k_3=0}^{n-s} (-1)^{k_1+k_2} \binom{r_1-1}{k_1} \binom{r_2-1}{k_2} \binom{n-s}{k_3} \frac{B(A_1, A_2) \Gamma(a_1 + m_1) \Gamma(a_2 + m_2 - 1/2)}{2\{a_1/b_1 + \Omega(x_{1j})\}^{(a_1+m_1)} 2\{a_2/b_2 + \Omega(x_{2j})\}^{(a_2+m_2-1/2)}} - 1 \right]$$

$$\rho(\hat{p}_{(KLF)}) = 2 \left[N^{-2} \sum_{k_1=0}^{r_1-1} \sum_{k_2=0}^{r_2-1} \sum_{k_3=0}^{n-s} (-1)^{k_1+k_2} \binom{r_1-1}{k_1} \binom{r_2-1}{k_2} \binom{n-s}{k_3} \frac{B(A_1 + 1, A_2) \Gamma(a_1 + m_1) \Gamma(a_2 + m_2)}{2\{a_1/b_1 + \Omega(x_{1j})\}^{(a_1+m_1)} 2\{a_2/b_2 + \Omega(x_{2j})\}^{(a_2+m_2)}} \right. \\ \left. \times \sum_{k_1=0}^{r_1-1} \sum_{k_2=0}^{r_2-1} \sum_{k_3=0}^{n-s} (-1)^{k_1+k_2} \binom{r_1-1}{k_1} \binom{r_2-1}{k_2} \binom{n-s}{k_3} \frac{B(A_1 - 1, A_2) \Gamma(a_1 + m_1) \Gamma(a_2 + m_2)}{2\{a_1/b_1 + \Omega(x_{1j})\}^{(a_1+m_1)} 2\{a_2/b_2 + \Omega(x_{2j})\}^{(a_2+m_2)}} - 1 \right]$$

Where N^{-1} is formulized as

$$N^{-1} = \sum_{k_1=0}^{r_1-1} \sum_{k_2=0}^{r_2-1} \sum_{k_3=0}^{n-s} (-1)^{k_1+k_2} \binom{r_1-1}{k_1} \binom{r_2-1}{k_2} \binom{n-s}{k_3} \frac{B(A_1, A_2) \Gamma(a_1 + m_1) \Gamma(a_2 + m_2)}{2\{a_1/b_1 + \Omega(x_{1j})\}^{(a_1+m_1)} 2\{a_2/b_2 + \Omega(x_{2j})\}^{(a_2+m_2)}}$$

$$A_1 = n - s - k_3 + s_1 + c_1 \text{ and } A_2 = s_2 + k_3 + d_1$$

Similarly, expressions for Bayes estimators and their posterior risks under the rest of the priors can be obtained with little modifications.

Elicitation

In Bayesian analysis the elicitation of opinion is a crucial step. In statistical inference, the characteristics of a certain predictive distribution proposed by an expert determine the hyper-parameters of a prior distribution. Focus on a method of elicitation based on prior predictive distribution. The elicitation of hyper-parameter from the prior $p(\lambda)$ is a difficult task. The prior predictive distribution is used for the elicitation of the hyper-parameters which is compared with the experts' judgment about this distribution and then the hyper-parameters are chosen in such a way so as to make the judgment agree as closely as possible with the given distribution. See also Grimshaw et al. (2001), O'Hagan et al. (2006), Jenkinson (2005) and Leon et al. (2003). According to Aslam (2003), the method of elicitation is to compare the prior predictive distribution with experts' assessment about this distribution and then to choose the hyper-parameters that make the assessment agree closely with the member of the family. The prior predictive distributions under all the priors are derived using:

$$p(y) = \int_{\Theta} p(y | \Theta) p(\Theta) d\Theta$$

Elicitation under Nakagami distribution

The prior predictive distribution using Nakagami prior is:

$$p(y) = 2(a_1 b_1^{-1})^{a_1} \frac{y a_1 c_1}{(c_1 + d_1)(y^2 + a_1 b_1^{-1})^{(a_1+1)}} + 2(a_2 b_2^{-1})^{a_2} \frac{y a_2 d_1}{(c_1 + d_1)(y^2 + a_2 b_2^{-1})^{(a_2+1)}}, y > 0 \quad (19)$$

For the elicitation of the six hyper-parameters, six different intervals are considered. From Equation (19), the experts' probabilities/assessments are supposed to be 0.10 for each case. The six integrals for equation (19) are considered with the following limits of the values of random variable 'Y': (0, 10), (10, 20), (20, 30), (30, 40), (40, 50) and (50, 60) respectively. For the elicitation of the hyper-parameters a_1, a_2, b_1, b_2, c_1 , and d_1 . These six equations are solved simultaneously through computer program developed in *SAS* package using the command of PROC SYSLIN. Thus

the values of hyper-parameters obtained by applying this methodology are: 0.000231, 0.012109, 0.52114, 4.99325, 0.52130, and 0.14790 respectively.

Elicitation under Chi Prior

The prior predictive distribution using Chi prior is:

$$p(y) = \frac{(0.5)^{\frac{e_1}{2}} y e_1 c_2}{(c_2 + d_2)(y^2 + 0.50)^{(e_1/2+1)}} + \frac{(0.5)^{\frac{e_2}{2}} y e_2 d_2}{(c_2 + d_2)(y^2 + 0.50)^{(e_2/2+1)}}, y > 0$$

Now, elicit four hyper-parameters, so consider the four integrals. The expert probabilities are assumed to 0.15 for each integral with the following limits of the values of random variable 'Y': (0, 15), (15, 30), (30, 45) and (45, 60). Using the similar kind of program, as discussed above, we have the following values of the hyper-parameters $e_1 = 20.1056$, $e_2 = 14.23569$, $c_2 = 0.09377$ and $d_2 = 0.08749$.

Elicitation under Rayleigh Prior

The prior predictive distribution using Rayleigh prior is:

$$p(y) = \frac{(v_1)^{-2} y c_3}{(c_3 + d_3)(y^2 + 2v_1^{-2})^2} + \frac{(v_2)^{-2} y d_3}{(c_3 + d_3)(y^2 + 2v_2^{-2})^2}, y > 0$$

Again, elicit four hyper-parameters, so consider the four integrals. The expert probabilities are assumed to 0.15 for each integral with the following limits of the values of random variable 'Y': (0, 15), (15, 30), (30, 45) and (45, 60). Using the similar kind of program, as discussed above, we attained the following values of the hyper-parameters $v_1 = 5.052104$, $v_2 = 5.03251$, $c_3 = 0.67213$ and $d_3 = 0.91035$.

Simulation Study and Comparisons

A simulation study is carried out in order to investigate the performance of Bayes estimators under tenfold choice of the parametric values, different sample sizes, and the different values of the mixing proportion. Take random samples of sizes $n = 20, 40$, and 80 from the two component mixture of Rayleigh distributions with tenfold choice of parameters

BAYESIAN ESTIMATION OF TWO-COMPONENT MIXTURE

$(\lambda_1, \lambda_2) \in \{(0.1, 0.12), (1, 1.2), (10, 12), (0.1, 12), (10, 0.12)\}$, $p_1 = 0.45$ and 0.6 .

To generate a mixture data we make use of probabilistic mixing with probabilities p_1 and $(1 - p_1)$. A uniform number u is generated n times and if $u < p_1$ the observation is taken randomly from F_1 (the Rayleigh distribution with parameter λ_1) otherwise from F_2 (from the Rayleigh distribution with parameter λ_2). The choice of the censoring time is made in such a way that the censoring rate in the resultant sample is approximately 20%. To implement censored samplings, we considered that the $x_{1r_1}, \dots, x_{1s_1}$ and $x_{2r_2}, \dots, x_{2s_2}$ failed items come from first and second subpopulations respectively. The rest of the observations which are less than x_r and greater than x_s have been assumed to be censored from each component.

Where $x_s = \max(x_{1s_1}, x_{2s_2})$ and $x_r = \min(x_{1r_1}, x_{2r_2})$. The simulated data sets have been obtained using following steps:

- Step 1:** Draw samples of size 'n' from the mixture model
- Step 2:** Generate a uniform random no. u for each observation
- Step 3:** If $u \leq \pi$, the take the observation from first subpopulation otherwise from the second subpopulation
- Step 4:** Determine the test termination points on left and right, that is, determine the values of x_r and x_s
- Step 5:** The observations which are less than x_r and greater than x_s have been considered to be censored from each component
- Step 6:** Use the remaining observations from each component for the analysis

To avoid an extreme sample, we simulate 10,000 data sets each of size n . The Bayes estimates and posterior risks (in parenthesis) are computed using Mathematica 8.0. The average of these estimates and corresponding risks are reported in tables 1- 15. The abbreviations used in the tables are: B.Es: Bayes estimators; P.Rs: Posterior risks; NP: Nakagami prior; CP: Chi prior; RP: Rayleigh prior.

Table 1: B.Es and P.Rs under NP using $(\lambda_1, \lambda_2, p_1) = (0.1, 0.12, 0.45)$ and $(0.1, 0.12, 0.6)$

n	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}_1	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}_1
squared error loss function						
20	0.104076 (0.000479)	0.127713 (0.000558)	0.498425 (0.013229)	0.099628 (0.000285)	0.135951 (0.000904)	0.665779 (0.011667)
40	0.099427 (0.000223)	0.12652 (0.000306)	0.48622 (0.007231)	0.094406 (0.000127)	0.131125 (0.000431)	0.659375 (0.006346)
80	0.099036 (0.000114)	0.125807 (0.000161)	0.478841 (0.003865)	0.092618 (0.000057)	0.13063 (0.000230)	0.61648 (0.003218)
k-loss function						
20	0.101884 (0.086648)	0.123181 (0.069120)	0.480102 (0.129905)	0.095373 (0.056005)	0.136428 (0.104087)	0.655312 (0.061058)
40	0.101669 (0.045012)	0.123008 (0.037679)	0.471869 (0.070063)	0.096164 (0.027334)	0.131118 (0.053150)	0.645074 (0.031642)
80	0.090768 (0.021446)	0.121778 (0.019760)	0.470942 (0.034345)	0.097312 (0.014017)	0.125869 (0.028207)	0.640866 (0.016573)

Table 2: B.Es and P.Rs under NP using $(\lambda_1, \lambda_2, p_1) = (1, 1.2, 0.45)$ and $(1, 1.2, 0.60)$

n	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}_1	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}_1
squared error loss function						
20	1.03790 (0.046897)	1.26375 (0.054602)	0.498181 (0.013210)	0.978002 (0.027554)	1.35085 (0.087625)	0.665995 (0.011594)
40	1.00642 (0.022787)	1.25934 (0.028565)	0.482283 (0.007234)	0.979230 (0.013363)	1.31290 (0.045048)	0.657711 (0.005650)
80	0.996073 (0.011180)	1.25518 (0.015747)	0.478649 (0.003863)	0.989340 (0.006855)	1.307418 (0.023781)	0.616586 (0.003222)
k-loss function						
20	1.02547 (0.085012)	1.297250 (0.069721)	0.484873 (0.126936)	10.11040 (0.083902)	12.737600 (0.068839)	0.481255 (0.129790)
40	0.972684 (0.043184)	1.24985 (0.037322)	0.477066 (0.068486)	9.85076 (0.043278)	12.45580 (0.037003)	0.474996 (0.069297)
80	0.994972 (0.024387)	1.22715 (0.021374)	0.469378 (0.037459)	9.91883 (0.022678)	12.11990 (0.019891)	0.468498 (0.036650)

BAYESIAN ESTIMATION OF TWO-COMPONENT MIXTURE

Table 3: B.Es and P.Rs under NP using $(\lambda_1, \lambda_2, p_1) = (10, 12, 0.45)$ and $(10, 12, 0.60)$

n	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}_1	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}_1
squared error loss function						
20	10.82959 (4.256710)	12.9605 (4.8388)	0.497084 (0.013244)	9.58255 (2.66321)	13.61630 (7.85146)	0.663355 (0.01170)
40	10.12890 (2.39442)	12.88640 (2.71684)	0.479166 (0.007305)	9.77352 (1.40887)	13.49536 (4.29918)	0.656693 (0.006435)
80	9.61493 (1.05203)	12.67810 (1.58376)	0.462094 (0.003820)	9.88275 (0.58717)	13.45520 (2.30094)	0.616606 (0.003233)
k-loss function						
20	10.11040 (0.083902)	12.73760 (0.068839)	0.481255 (0.12979)	9.84880 (0.056169)	12.80650 (0.104837)	0.653839 (0.061849)
40	9.85076 (0.043278)	12.45580 (0.037003)	0.474996 (0.069297)	9.94419 (0.027645)	12.63070 (0.054150)	0.651788 (0.031611)
80	9.91883 (0.022678)	12.11990 (0.019891)	0.468498 (0.036650)	9.95821 (0.013426)	12.58110 (0.028228)	0.640724 (0.015684)

Table 4: B.Es and P.Rs under NP using $(\lambda_1, \lambda_2, p_1) = (0.10, 12, 0.45)$ and $(0.10, 12, 0.60)$

n	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}_1	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}_1
squared error loss function						
20	0.095619 (0.000301)	13.67980 (4.868910)	0.534912 (0.012036)	0.092487 (0.000201)	13.81620 (7.189420)	0.687435 (0.010396)
40	0.090655 (0.000131)	13.54530 (2.343030)	0.51823 (0.006457)	0.091992 (0.000096)	13.63860 (3.41382)	0.677511 (0.005650)
80	0.090905 (0.000065)	13.48370 (1.148460)	0.509322 (0.003346)	0.091260 (0.000048)	13.5980 (1.767790)	0.672210 (0.002953)
k-loss function						
20	0.0914225 (0.067737)	12.87110 (0.053268)	0.522308 (0.097694)	0.093144 (0.048171)	13.86210 (0.078296)	0.679012 (0.049925)
40	0.092511 (0.032515)	12.72830 (0.025796)	0.511618 (0.052023)	0.0951075 (0.023389)	13.70450 (0.037372)	0.673099 (0.026303)
80	0.09455 (0.015937)	12.37891 (0.012696)	0.505934 (0.026871)	0.096113 (0.011527)	12.96763 (0.018268)	0.669953 (0.013511)

Table 5: B.Es and P.Rs under NP using $(\lambda_1, \lambda_2, p_1) = (10, 0.12, 0.45)$ and $(10, 0.12, 0.60)$

n	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}_1	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}_1
squared error loss function						
20	11.92930 (4.278770)	0.112064 (0.000325)	0.403231 (0.011879)	11.53870 (3.16833)	0.113048 (0.000483)	0.585750 (0.011740)
40	11.5720 (2.13578)	0.114125 (0.000155)	0.412042 (0.006265)	11.46580 (1.51987)	0.11513 (0.000219)	0.586132 (0.006334)
80	11.17240 (1.08183)	0.118567 (0.000075)	0.42970 (0.003216)	11.40931 (0.756977)	0.116862 (0.000103)	0.589618 (0.003294)
k-loss function						
20	11.64350 (0.067737)	0.10762 (0.053254)	0.417776 (0.15071)	10.98420 (0.048171)	0.106840 (0.078242)	0.574552 (0.078744)
40	11.44770 (0.032519)	0.118174 (0.025790)	0.423946 (0.080978)	10.87421 (0.023391)	0.108845 (0.037356)	0.586549 (0.041779)
80	11.15780 (0.015944)	0.119735 (0.012698)	0.438659 (0.042023)	10.74670 (0.011528)	0.109872 (0.018256)	0.596061 (0.021540)

Table 6: B.Es and P.Rs under CP using $(\lambda_1, \lambda_2, p_1) = (0.1, 0.12, 0.45)$ and $(0.1, 0.12, 0.60)$

n	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}_1	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}_1
squared error loss function						
20	0.160528 (0.000519)	0.169018 (0.000592)	0.479278 (0.013082)	0.134094 (0.000277)	0.195013 (0.000893)	0.664322 (0.0116345)
40	0.133714 (0.000286)	0.147046 (0.000329)	0.468172 (0.007333)	0.11004 (0.000116)	0.172682 (0.000448)	0.663936 (0.006115)
80	0.111222 (0.000147)	0.139219 (0.000196)	0.448705 (0.004006)	0.103352 (0.000058)	0.151533 (0.000223)	0.66273 (0.003196)
k-loss function						
20	0.161522 (0.04053)	0.158551 (0.043568)	0.465088 (0.145568)	0.13577 (0.029991)	0.198858 (0.049799)	0.653999 (0.061370)
40	0.122886 (0.030112)	0.14806 (0.031199)	0.464529 (0.072191)	0.112829 (0.019077)	0.167765 (0.034061)	0.652043 (0.030841)
80	0.104790 (0.02132)	0.132506 (0.020869)	0.463772 (0.038239)	0.101234 (0.010604)	0.143011 (0.020665)	0.650817 (0.015416)

BAYESIAN ESTIMATION OF TWO-COMPONENT MIXTURE

Table 7: B.Es and P.Rs under CP using $(\lambda_1, \lambda_2, p_1) = (1, 1.2, 0.45)$ and $(1, 1.2, 0.60)$

n	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}_1	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}_1
squared error loss function						
20	1.52970 (0.046080)	1.57752 (0.051958)	0.476725 (0.013179)	1.35361 (0.027818)	1.88630 (0.080777)	0.663527 (0.011650)
40	1.26130 (0.026077)	1.41822 (0.031378)	0.469789 (0.007406)	1.16501 (0.013935)	1.66429 (0.044256)	0.659758 (0.006292)
80	1.14492 (0.014751)	1.25617 (0.015752)	0.45906 (0.004146)	1.05707 (0.006667)	1.42491 (0.021669)	0.656866 (0.003355)
k-loss function						
20	1.53236 (0.041669)	1.49543 (0.045185)	0.454182 (0.147876)	1.32818 (0.028990)	1.84131 (0.048999)	0.654349 (0.0615537)
40	1.25475 (0.028827)	1.43835 (0.028914)	0.452576 (0.073808)	1.13137 (0.019386)	1.63107 (0.034167)	0.654126 (0.031362)
80	1.06717 (0.018469)	1.33081 (0.018471)	0.450778 (0.036839)	1.04654 (0.0119729)	1.43309 (0.024099)	0.653583 (0.016246)

Table 8: B.Es and P.Rs under CP using $(\lambda_1, \lambda_2, p_1) = (10, 12, 0.45)$ and $(10, 12, 0.60)$

n	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}_1	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}_1
squared error loss function						
20	5.53623 (0.433161)	5.39552 (0.43394)	0.470341 (0.013715)	5.58208 (0.415167)	5.00538 (0.452878)	0.640696 (0.012735)
40	6.27422 (0.39050)	6.44220 (0.396637)	0.456349 (0.0072919)	6.62719 (0.356135)	5.85830 (0.422084)	0.629771 (0.006889)
80	7.34364 (0.339835)	7.76371 (0.337258)	0.45011 (0.003809)	7.60293 (0.287163)	7.02488 (0.384135)	0.625043 (0.003610)
k-loss function						
20	5.52495 (0.029004)	5.37683 (0.030637)	0.453655 (0.14971)	5.8203 (0.024786)	4.95789 (0.037429)	0.629618 (0.072137)
40	6.24029 (0.020289)	6.40839 (0.019526)	0.447779 (0.077333)	6.59809 (0.016483)	5.99867 (0.025214)	0.623712 (0.037649)
80	7.05432 (0.012209)	7.72934 (0.010219)	0.448146 (0.040194)	7.46259 (0.010121)	6.96302 (0.015946)	0.622423 (0.019236)

Table 9: B.Es and P.Rs under CP using $(\lambda_1, \lambda_2, p_1) = (0.10, 12, 0.45)$ and $(0.10, 12, 0.60)$

n	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}_1	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}_1
squared error loss function						
20	0.135859 (0.000265)	5.51759 (0.447943)	0.526231 (0.012354)	0.13051 (0.000208)	5.08452 (0.461858)	0.682633 (0.010735)
40	0.115993 (0.000132)	6.65447 (0.40904)	0.513532 (0.006542)	0.112149 (0.000099)	6.04774 (0.435674)	0.674904 (0.005747)
80	0.099126 (0.000059)	8.88344 (0.346268)	0.506878 (0.003368)	0.102298 (0.000048)	7.33437 (0.384372)	0.670852 (0.002976)
k-loss function						
20	0.142139 (0.028685)	5.48345 (0.030311)	0.513036 (0.104196)	0.129112 (0.024474)	5.04152 (0.037048)	0.673849 (0.052484)
40	0.115279 (0.019664)	6.60253 (0.018872)	0.506765 (0.053772)	0.110234 (0.015909)	6.371890 (0.024397)	0.670396 (0.026986)
80	0.105346 (0.012065)	8.09129 (0.010753)	0.503448 (0.027325)	0.104518 (0.009303)	7.380910 (0.015178)	0.668463 (0.0141179)

Table 10: B.Es and P.Rs under CP using $(\lambda_1, \lambda_2, p_1) = (10, 0.12, 0.45)$ and $(10, 0.12, 0.60)$

n	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}_1	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}_1
squared error loss function						
20	5.62845 (0.441878)	0.146896 (0.000328)	0.421965 (0.012086)	5.96024 (0.424476)	0.151952 (0.000431)	0.578365 (0.012083)
40	6.44104 (0.400232)	0.127722 (0.000153)	0.43595 (0.006316)	6.85842 (0.368546)	0.136448 (0.000225)	0.578932 (0.006429)
80	7.58808 (0.343598)	0.117307 (0.000074)	0.45587 (0.0032287)	8.06977 (0.302239)	0.120046 (0.000104)	0.5886153 (0.003318)
k-loss function						
20	5.58490 (0.028685)	0.14275 (0.030323)	0.405755 (0.162971)	5.89857 (0.024474)	0.159564 (0.037102)	0.566651 (0.083544)
40	6.40615 (0.019662)	0.126893 (0.018875)	0.417657 (0.084299)	6.84141 (0.015908)	0.132547 (0.024410)	0.561311 (0.043066)
80	7.44424 (0.01207)	0.117069 (0.010756)	0.439330 (0.042888)	7.71344 (0.009358)	0.121066 (0.014498)	0.558488 (0.021872)

BAYESIAN ESTIMATION OF TWO-COMPONENT MIXTURE

Table 11: B.Es and P.Rs under RP using $(\lambda_1, \lambda_2, p_1) = (0.1, 0.12, 0.45)$ and $(0.1, 0.12, 0.60)$

n	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}_1	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}_1
squared error loss function						
20	0.105782 (0.000427)	0.135123 (0.000581)	0.434012 (0.012329)	0.103079 (0.000282)	0.138375 (0.000938)	0.645401 (0.003282)
40	0.10453 (0.000233)	0.131766 (0.000296)	0.474736 (0.007004)	0.096193 (0.000128)	0.138033 (0.000453)	0.641649 (0.006242)
80	0.096183 (0.000103)	0.129324 (0.0001600)	0.468102 (0.003756)	0.093818 (0.000058)	0.130254 (0.000234)	0.64095 (0.003206)
k-loss function						
20	0.101651 (0.0687089)	0.123051 (0.066684)	0.485036 (0.119068)	0.093688 (0.049411)	0.114119 (0.088853)	0.630198 (0.06578)
40	0.101989 (0.042361)	0.112941 (0.036825)	0.467163 (0.069873)	0.0975664 (0.026666)	0.129533 (0.049849)	0.624327 (0.031978)
80	0.100056 (0.022779)	0.118725 (0.020242)	0.46499 (0.037457)	0.0978534 (0.0121946)	0.124279 (0.023696)	0.623173 (0.014910)

Table 12: B.Es and P.Rs under RP using $(\lambda_1, \lambda_2, p_1) = (1, 1.2, 0.45)$ and $(1, 1.2, 0.60)$

n	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}_1	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}_1
squared error loss function						
20	1.05922 (0.042667)	1.33314 (0.056026)	0.433188 (0.012338)	1.03293 (0.027869)	1.51575 (0.098507)	0.644157 (0.011345)
40	1.00664 (0.021771)	1.29054 (0.029663)	0.476983 (0.007093)	0.970745 (0.012762)	1.39351 (0.045143)	0.642962 (0.006225)
80	0.983895 (0.011111)	1.289237 (0.016032)	0.475851 (0.003774)	0.983541 (0.011043)	1.37069 (0.019325)	0.62145 (0.003245)
k-loss function						
20	1.16454 (0.072345)	1.28713 (0.063822)	0.468544 (0.131557)	0.996425 (0.064813)	1.09914 (0.119828)	0.62148 (0.070912)
40	1.00749 (0.041506)	1.12108 (0.035513)	0.466946 (0.069635)	0.985194 (0.027093)	1.35608 (0.050711)	0.620150 (0.032581)
80	0.954177 (0.020998)	1.11020 (0.019217)	0.457543 (0.034475)	0.995916 (0.013199)	1.34535 (0.025414)	0.618873 (0.015959)

Table 13: B.Es and P.Rs under RP using $(\lambda_1, \lambda_2, p_1) = (10, 12, 0.45)$ and $(10, 12, 0.60)$

n	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}_1	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}_1
squared error loss function						
20	9.58324	11.43070	0.427949	9.58708	10.9565	0.634954
	(3.169330)	(3.83663)	(0.012469)	(2.322100)	(5.00616)	(0.011817)
40	9.67622	11.78300	0.47243	9.599110	11.9233	0.630872
	(1.924840)	(2.34378)	(0.007076)	(1.29289)	(3.42684)	(0.006453)
80	9.68654	12.34210	0.470593	9.73014	12.86730	0.615151
	(0.947672)	(1.416400)	(0.003749)	(0.586285)	(2.08719)	(0.003283)
k-loss function						
20	8.12011	11.32660	0.464824	10.09989	12.1777	0.629444
	(0.068480)	(0.058682)	(0.133263)	(0.0493049)	(0.088018)	(0.066076)
40	9.37871	11.84870	0.461711	9.41742	11.96030	0.616346
	(0.038674)	(0.034311)	(0.069343)	(0.027143)	(0.051405)	(0.034086)
80	9.42150	11.9648	0.460929	9.571406	12.45630	0.610353
	(0.021294)	(0.019479)	(0.036338)	(0.012793)	(0.025410)	(0.015892)

Table 14: B.Es and P.Rs under RP using $(\lambda_1, \lambda_2, p_1) = (0.10, 12, 0.45)$ and $(0.10, 12, 0.60)$

n	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}_1	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}_1
squared error loss function						
20	0.094884	11.89990	0.469921	0.098341	11.96510	0.664261
	(0.000263)	(3.293650)	(0.011542)	(0.000209)	(4.61595)	(0.010333)
40	0.095518	12.63033	0.509872	0.098953	12.78527	0.656383
	(0.000137)	(1.93028)	(0.006313)	(0.000098)	(2.78527)	(0.005624)
80	0.091143	12.39790	0.495107	0.099681	12.49280	0.646003
	(0.000063)	(1.10531)	(0.003307)	(0.000047)	(1.58923)	(0.002943)
k-loss function						
20	0.114775	12.94880	0.506062	0.107666	13.3107	0.655632
	(0.059662)	(0.048205)	(0.099563)	(0.043940)	(0.067830)	(0.052989)
40	0.094401	12.89850	0.503309	0.091838	12.75870	0.646092
	(0.030531)	(0.024543)	(0.052498)	(0.022345)	(0.034796)	(0.027125)
80	0.091618	12.72390	0.501733	0.0904158	12.41740	0.636729
	(0.015444)	(0.012387)	(0.026989)	(0.011268)	(0.017632)	(0.013727)

BAYESIAN ESTIMATION OF TWO-COMPONENT MIXTURE

Table 15: B.Es and P.Rs under RP using $(\lambda_1, \lambda_2, p_1) = (10, 0.12, 0.45)$ and $(10, 0.12, 0.60)$

n	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}_1	$\hat{\lambda}_1$	$\hat{\lambda}_2$	\hat{p}_1
squared error loss function						
20	10.7884 (3.33464)	0.11846 (0.000331)	0.372751 (0.010833)	10.50830 (2.35803)	0.119567 (0.000474)	0.567091 (0.0113719)
40	10.3168 (1.92637)	0.11824 (0.000153)	0.406198 (0.006094)	10.41890 (1.30849)	0.116756 (0.000221)	0.567909 (0.006219)
80	10.30871 (0.992713)	0.119613 (0.000074)	0.427844 (0.003169)	10.39070 (0.686735)	0.1198576 (0.0001029)	0.585739 (0.003262)
k-loss function						
20	9.81163 (0.059662)	0.152163 (0.048174)	0.406292 (0.150848)	9.22044 (0.043940)	0.127570 (0.067755)	0.555927 (0.081129)
40	10.94770 (0.030533)	0.111752 (0.024540)	0.419822 (0.080943)	10.64808 (0.022345)	0.12357 (0.034789)	0.55487 (0.042404)
80	10.57770 (0.015446)	0.112851 (0.012385)	0.429373 (0.042003)	10.41660 (0.011269)	0.123356 (0.017625)	0.550732 (0.021699)

Numerical results of the simulation study, presented in tables 1-15, reveal interesting properties of the proposed Bayes estimators. The estimated values of the parameters converge to the true values, and amounts of posterior risks tend to decrease for larger choice of sample size. Another interesting point concerning the posterior risks of the estimates of λ_1, λ_2 is that increasing (decreasing) the proportion of the component in mixture reduces (increases) the amount of the posterior risk for the estimates of λ_1 . In addition, when SELF is assumed and values of λ_i are relatively smaller i.e. for $(\lambda_1, \lambda_2) = (0.1, 0.12)$ and $(1, 1.2)$, the Bayes estimates assuming Rayleigh prior are more precise than the rest of the informative priors, as the averaged posterior risks of the mixture components are smaller as compared to those under other priors. On the other hand, for quite larger values of parameters, i.e. for $(\lambda_1, \lambda_2) = (10, 12)$, and for significantly different values of the parameters, i.e. for $(\lambda_1, \lambda_2) = (0.1, 12)$, the estimates under chi prior (with few exceptions) perform better than those under Nakagami and Rayleigh priors. However, the estimates for the mixing parameter (p_1), under Rayleigh prior, are associated with the minimum amounts of posterior risks irrespective of choice of true parametric values. When KLF is assumed, the estimates under chi prior are found to be the most efficient for all combinations of the values of the parameters, with an exception in case of $(\lambda_1, \lambda_2) = (0.1, 0.12)$, where the estimates under the assumption of Nakagami prior are better than those under other priors. However,

the estimates for the mixing parameter (p_1) are having mixed behavior, as for various choices of the true parametric values indicate the preference of different priors.

The Bayes estimates of the lifetime parameters are over/under-estimated but the size of over/under-estimation is greater under squared error loss function. On the other hand, estimates of the mixing proportion parameter have mixed behavior sometimes over-estimated and sometimes under-estimated, but the Bayes estimates under Rayleigh prior are much closer to the true parametric value. In comparison of loss functions it has been assessed that the magnitudes of posterior risks under squared error loss function are smaller than those under k-loss function for smaller choice of true parametric values i.e. for $(\lambda_1, \lambda_2) = (0.1, 0.12)$ and $(1, 1.2)$. On the other hand, for quite larger values of parameters, i.e. for $(\lambda_1, \lambda_2) = (10, 12)$, and for significantly different values of the parameters, i.e. for $(\lambda_1, \lambda_2) = (0.1, 12)$ and $(10, 0.12)$, the k-loss function produces the better results. It should also be mentioned here that the squared error loss function produces better convergence than k-loss function.

Real Data Analysis

In this section, real datasets are analyzed to illustrate the methodology discussed in the previous sections. In order to show the usefulness of the proposed mixture model, we applied the findings of the paper to the survival times (in years) of a group of patients given chemotherapy treatment. The data has been reported by Bekker et al. (2000). We have used the Kolmogorov-Smirnov and chi square tests to see whether the data follow the Rayleigh distribution. These tests say that the data follow the Rayleigh distribution at 5% level of significance with p-values 0.2170 and 0.2681 respectively. The data consisting of 46 survival times (in years) for 46 patients are:

Table 16: Survival times (in years) of patients given chemotherapy treatment

0.047, 0.115, 0.121, 0.132, 0.164, 0.197, 0.203, 0.260, 0.282, 0.296, 0.334, 0.395,
0.458, 0.466, 0.501, 0.507, 0.529, 0.534, 0.540, 0.570, 0.641, 0.644, 0.696, 0.841,
0.863, 1.099, 1.219, 1.271, 1.326, 1.447, 1.485, 1.553, 1.581, 1.589, 2.178, 2.343,
2.416, 2.444, 2.825, 2.830, 3.578, 3.658, 3.743, 3.978, 4.003, 4.033.

BAYESIAN ESTIMATION OF TWO-COMPONENT MIXTURE

Consider the case when the data are doubly type II censored. Data are randomly grouped into two sets using probabilistic mixing for $p_1 = 0.60$.

Table 17: Doubly censored mixture real life data regarding survival times of patients given chemotherapy treatment

Population-I	Population-II
0.197, 0.534, 0.115, 0.296, 0.121, 0.466, 0.529, 1.447, 0.863, 0.132, 0.395, 0.696, 2.825, 3.658, 3.978, 3.743, 2.343, 2.178, 0.540, 4.003, 1.553, 1.485, 2.83, 2.416	0.260, 1.099, 0.501, 0.458, 0.641, 0.334, 0.570, 0.164, 0.203, 0.282, 0.047, 1.271, 1.589, 1.326, 0.841, 2.444

The following characteristics are extracted from censored data for the analysis of mixture model:

$$p_1 = 0.6, n = 40, r = 5, r_1 = 2, r_2 = 3, n - r = 9, s = 36, s_1 = 22, s_2 = 14, n_1 = 24, n_2 = 16, x_{r_1} = 0.121, x_{s_1} = 3.978, x_{r_2} = 0.203, \text{ and } x_{s_2} = 2.444.$$

$$\sum_{i=r_1}^{s_1} x_{1(i)}^2 = 84.6037 \text{ and } \sum_{i=r_2}^{s_2} x_{2(i)}^2 = 15.2833.$$

The similar methodology has been employed when $p_1 = 0.45$.

$$p_1 = 0.45, n = 40, r = 5, r_1 = 2, r_2 = 3, n - r = 9, s = 36, s_1 = 16, s_2 = 20, n_1 = 18, n_2 = 22, x_{r_1} = 0.121, x_{s_1} = 3.658, x_{r_2} = 0.164, \text{ and } x_{s_2} = 3.978,$$

$$\sum_{i=r_1}^{s_1} x_{1(i)}^2 = 48.704 \text{ and } \sum_{i=r_2}^{s_2} x_{2(i)}^2 = 37.1999.$$

Bayes estimates are obtained assuming informative priors under squared error loss function, and k-loss function.

Table 18: B.Es and P.Rs in parentheses under squared error loss function, and k-loss function for real data set.

Priors	squared error loss function			k-loss function		
	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\rho}_1$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\rho}_1$
p1 = 0.6						
Nakagami Prior	0.383514 (0.001682)	0.947312 (0.016257)	0.677477 (0.005652)	0.38129 (0.023399)	0.938511 (0.037685)	0.673064 (0.026312)
Chi Prior	0.462925 (0.001678)	1.148390 (0.015708)	0.674903 (0.005747)	0.461094 (0.015909)	1.141450 (0.024405)	0.670395 (0.026986)
Rayleigh Prior	0.392190 (0.001681)	0.980312 (0.016213)	0.665364 (0.005626)	0.390017 (0.022350)	0.971852 (0.034973)	0.6608996 (0.027130)
p1 = 0.45						
Nakagami Prior	0.377204 (0.0024201)	0.722023 (0.007572)	0.516536 (0.006537)	0.373998 (0.034429)	0.716491 (0.030999)	0.509802 (0.053186)
Chi Prior	0.482646 (0.002974)	0.837202 (0.008188)	0.511665 (0.006693)	0.479778 (0.023980)	0.831775 (0.026185)	0.504667 (0.055851)
Rayleigh Prior	0.388672 (0.002412)	0.740606 (0.007504)	0.508438 (0.006381)	0.385574 (0.032264)	0.735272 (0.029121)	0.50177 (0.053512)

The findings from the real life analysis are in close accordance with those of simulation study. It can be assessed that the chi prior produces better results for parameters λ_1 and λ_2 , while in case of mixing parameter the Rayleigh prior provides comparatively better results than other priors. It should further be noted that the estimates under squared error loss function are associated with smaller amounts of posterior risks.

Conclusion

The Bayesian inference of the mixture of Rayleigh model under doubly type II censoring has been considered assuming informative priors. The simulation study has displayed some interesting properties of the Bayes estimates. It is noted in each case that the posterior risks of estimates of lifetime parameters are reduced as the sample size increases. The results indicated that by using SELF and relatively smaller values of λ_i i.e. for $(\lambda_1, \lambda_2) = (0.1, 0.12)$ and $(1, 1.2)$, the Bayes estimates assuming Rayleigh prior are more precise than the rest of the informative priors. While, for quite larger values of parameters, i.e. for $(\lambda_1, \lambda_2) = (10, 12)$, and for significantly different values of the parameters, i.e. for $(\lambda_1, \lambda_2) = (0.1, 12)$ and $(10, 0.12)$, the estimates under chi prior perform better than other priors. Similarly, when KLF is considered, the estimates under chi prior are found to be the most efficient for most of the combinations of the values of the parameters. The performance of

the squared error loss function is better than k-loss function for $(\lambda_1, \lambda_2) = (0.1, 0.12)$ and $(1, 1.2)$. However, for $(\lambda_1, \lambda_2) = (10, 12)$, $(0.1, 12)$ and $(10, 0.12)$, the k-loss function produces the better results. It should also be mentioned here that the squared error loss function produces better convergence than k-loss function for almost all the cases. The real life example further strengthened the findings from the simulation study. The study can further be extended by considering some other censoring techniques, and using some more flexible probability distribution.

References

- Aslam, M. (2003). An application of prior predictive distribution to elicit the prior density. *Journal of Statistical Theory and Applications*, 2(1), 70-83.
- Bekker, A., Roux, J., & Mostert, P. (2000). A generalization of the compound Rayleigh distribution: using a Bayesian method on cancer survival times. *Communications in Statistics, Theory and Methods*, 29(7), 1419-1433.
- Blischke, W. R., & Murthy, D. N. P. (2000). *Reliability*. New York: John Wiley and Sons.
- Bolstad, W. M. (2004). *Introduction to Bayesian Statistics*. New York: John Wiley and Sons.
- Bluebaly, T., & Bougila, N. (2011). Bayesian learning of finite generalized Gaussian mixture models on images. *Signal Processing*, 91(4), 801-820.
- Fernandez, A. J. (2000). On maximum likelihood prediction based on type II doubly censored exponential data. *Metrika*, 50(3), 211-220.
- Fernandez, A. J. (2006). Bayesian estimation based on trimmed samples from Pareto populations. *Computational statistics & data analysis*, 51(2), 1119-1130.
- Ghosh, S. K., & Ebrahimi, N. (2001). *Bayesian analysis of the mixing function in a mixture of two exponential distributions*. Tech. Rep. 2531, Institute of Statistics Mimeographs, North Carolina State University, North Carolina State University.
- Grimshaw, S D., Collings, B. J., Larsen, W. A., & Hurt, C. R. (2001). Eliciting Factor Importance in a Designed Experiment. *Technometrics*, 43(2), 133-146.
- Jenkinson, D. (2005). *The elicitation of probabilities: A review of the statistical literature*. Department of Probability and Statistics, University of Sheffield.

- Khan, H. M., Albatineh, A., Alshahrani, S., Jenkins, N., & Ahmed, N. U. (2011). Sensitivity analysis of predictive modeling for responses from the three-parameter Weibull model with a follow-up doubly censored sample of cancer patients. *Computational Statistics & Data Analysis*, 55(12), 3093-3103.
- Khan, H. M., Provost, S. B., & Singh, A. (2010). Predictive inference from a two-parameter Rayleigh life model given a doubly censored sample. *Communications in Statistics—Theory and Methods*, 39(7), 1237-1246.
- Kim, C., & Song, S. (2010). Bayesian estimation of the parameters of the generalized exponential distribution from doubly censored samples. *Statistical Papers*, 51(3), 583-597.
- Kundu, D., & Howlader, H. (2010). Bayesian inference and prediction of the inverse Weibull distribution for type-II censored data. *Computational Statistics & Data Analysis*, 54(6), 1547-1558.
- Leon, J. C., Vazquez-Polo, J. F., & Gonzalez, L. R. (2003). Environmental and resource economics. *Kluwer Academic Publishers*, 26, 199-210.
- Mendenhall, W. & Hadar, R. J. (1958). Estimation of parameters of mixed exponentially distributed failure time distributions from censored life test data. *Biometrika*, 45, 504-520.
- Murthy, D. N. P., Bulmer, M. & Eccleston, J. A. (2004). Weibull model selection for reliability modelling. *Reliability Engineering & System Safety*, 86, 257-267.
- Noor, F. & Aslam, M. (2008). Bayesian inference of the inverse Weibull mixture distribution using type-I censoring. *Journal of Applied Statistics*, 40(5), 1076-1089.
- O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. E., Garthwaite, P. H., Jenkinson, D. J., ... Rakow, T. (2006). *Uncertain Judgements: Eliciting expert probabilities*. New York: John Wiley and Sons.
- Pak, A., Parham, G. A. & Saraj, M. (2013). On Estimation of Rayleigh Scale Parameter under Doubly Type-II Censoring from Imprecise Data. *Journal of Data Science*, 11, 305-322.
- Saleem, M., & Aslam, M. (2008). Bayesian analysis of the two component mixture of the Rayleigh distribution assuming the Uniform and the Jeffreys prior from censored data. *Journal of Applied Statistical Science*, 16(4), 105-113.
- Saleem, M., Aslam, M., & Economou, P. (2010). On the Bayesian analysis of the mixture of Power function distribution using the complete and the censored sample. *Journal of Applied Statistics*, 37(1), 25-40.

BAYESIAN ESTIMATION OF TWO-COMPONENT MIXTURE

Shi, Y., & Yan, W. (2010). The EB estimation of scale parameter for two parameter exponential distribution under the type-I censoring life test. *Journal of Physical Science*, 4, 25-30.

Singh, S. K., Singh, U., & Sharma, V. K. (2013). Bayesian analysis for Type-II hybrid censored sample from inverse Weibull distribution. *International Journal of System Assurance Engineering*, 1-8.

Soliman, A. A. (2006). Estimators for the finite mixture of Rayleigh model based on progressively censored data. *Communications in Statistics—Theory and Methods*, 35, 803-820.

Sultan, K. S., & Al-Moisheer, A. S. (2012). Approximate Bayes estimation of the parameters and reliability function of a mixture of two inverse Weibull distributions under type-2 censoring. *Journal of Statistical Computation and Simulation*, 1-15.

Sultan, K. S., Ismail, M. A., & Al-Moisheer, A. S. (2007). Mixture of two inverse Weibull distributions: Properties and estimation. *Computational Statistics & Data Analysis*, 51(11), 5377-5387.

Wasan, M. T. (1970). *Parametric estimation*. New York: McGraw-Hill.

Survival Estimation Using Bootstrap, Jackknife and K-Repeated Jackknife Methods

Adewara Johnson Ademola

University of Lagos
Akoka-Lagos, Nigeria

Ugochukwu Ahamefula Mbata

University of Lagos
Akoka-Lagos, Nigeria

Three re-sampling techniques are used to estimate the survival probabilities from an exponential life-time distribution. The aim is to employ a technique to obtain a parameter estimate for a two-parameter exponential distribution. The re-sampling methods considered are: Bootstrap estimation method (BE), Jackknife estimation method (JE) and the k -repeated Jackknife estimation method (KJE). The methods were computed to obtain the mean square error (MSE) and mean percentage error (MPE) based on simulated data. The estimates of the two-parameter exponential distribution were substituted to estimate survival probabilities. Results show that the MSE value is reduced when the K -repeated jackknife method is used.

Introduction

Modern statistics is anchored in the use of statistics and hypothesis tests that only have desirable and well-known properties when computed from populations that are normally distributed. While it is claimed that many such statistics and hypothesis tests are generally robust with respect to non-normality, other approaches that require an empirical investigation of the underlying population distribution or of the distribution of the statistic are possible and in some instances preferable. In instances when the distribution of a statistic, conceivably a very complicated statistic, is unknown, no recourse to a normal theory approach is available and alternative approaches are required. Statistical models and methods for estimating survival data and other time-to-event data are extensively used in many fields, including the biomedical sciences, engineering, the environmental sciences, economics, actuarial sciences, management, and the social sciences.

Adewara Johnson Ademola is a Senior Lecturer at the Distance Learning Institute at the University of Lagos. Email at jadewara@unilag.edu.ng

U.A. Mbata is a Graduate Fellow and Consulting Statistician in the Mathematics Department at the University of Lagos. Email at mmbataugochukwu@yahoo.com

SURVIVAL ESTIMATION USING THREE METHODS

Survival analysis refers to techniques for studying the occurrence and timing of events. It is concerned with studying the random variable T , representing the time between entry to a study and some event of interest, such as: death, the onset of disease, time until equipment failures, earthquakes, automobile accidents, time-to-promotions, time until stock market crashes, revolutions, job terminations, births, marriages, divorces, retirements or arrests. There are many different models for survival data, and what often distinguishes one model from another is the probability distribution for T . Resampling statistics refer to the use of the observed data or of a data generating mechanism (such as a die) to produce new hypothetical samples (resamples) that mimic an underlying population, the results of which can then be analyzed. With numerous cross-disciplinary applications especially in the sub-disciplines of the life science, resampling methods are widely used because they are options when parametric approaches are difficult to employ or otherwise do not apply.

Resampled data is derived using a manual mechanism to simulate many pseudo-trials. These approaches were difficult to utilize prior to 1980s because these methods require many repetitions. With the incorporation of computers, the trials can be simulated in a few minutes and is why these methods have become widely used. The methods that will be discussed are used to make many statistical inferences about the underlying population. The most practical use of resampling methods is to derive confidence intervals and test hypotheses. This is accomplished by drawing simulated samples from the data themselves (resamples) or from a reference distribution based on the data; afterwards, you are able to observe how the statistic of interest in these resamples behaves. Resampling approaches can be used to substitute for traditional statistical (formulaic) approaches or when a traditional approach is difficult to apply. These methods are widely used because their ease of use. They generally require minimal mathematical formulas, needing a small amount of mathematical (algebraic) knowledge. These methods are easy to understand and stray away from choosing an incorrect formula in your diagnostics.

Two general approaches considered here are: Jackknife approach and Bootstrap approach. The aim of this study is to employ a technique to obtain an estimate of the parameter of the two-parameter exponential distribution. The methods considered in this paper are: Bootstrap estimation method (BE), Jackknife estimation method (JE) and the k -repeated Jackknife estimation method (KJE). The estimates of the two-parameter exponential distributions are used to estimate the survival probability. Methodology under Bootstrap, Jackknife and the proposed K -repeated Jackknife is presented, followed by data analysis, results, discussion and a conclusion.

Two-Parameter Exponential Distribution

Re-sampling methods are becoming increasingly popular as statistical tools. These methods involve sampling or scrambling the original data numerous times. Two general approaches are considered here. They are: Jackknife approach and Bootstrap approach. The two-parameter exponential distribution is adopted when failure will never occur prior to some specified time, t_0 . The parameter t_0 is a location parameter that shifts the distribution an amount equal to t_0 towards the right on the time line. When $t \geq t_0$, the probability density function of exponential distribution becomes:

$$f(t; \theta) = \frac{1}{\theta} \exp\left(-\frac{1}{\theta}(t - t_0)\right), t \geq 0, \theta > 0, t \geq t_0 \quad (1)$$

and the survival function is given by:

$$S(t) = \int_t^{\infty} f(t; \theta) dt \quad (2)$$

where $\lambda = \frac{1}{\theta}$.

Bootstrap Estimation Method

Bootstrapping is a modern, computer-intensive, general purpose approach to statistical inference, falling within a broader class of re-sampling methods to simplify the often intricate calculations of traditional statistical theory. A parametric bootstrap method is considered in this article.

The general theory (see [Rizzo, 2008](#)) is as follows. Suppose t_1, t_2, \dots, t_n is a random sample from the distribution of T . An estimator $\hat{\theta}$ for a parameter θ is an n variate function $\hat{\theta} = \hat{\theta}(t_1, t_2, \dots, t_n)$ of the sample. Functions of the estimator $\hat{\theta}$ are therefore n -variate functions of the data, also. For simplicity, let $t = (t_1, t_2, \dots, t_n)^T \in R^n$, and $t^{(1)}, t^{(2)}, \dots$ denote a sequence of independent random samples generated from the distribution of T . Random variables from the sampling distribution of $\hat{\theta}$ can be generated by repeatedly drawing independent random samples $t^{(j)}$ and computing $\hat{\theta}^{(j)} = \hat{\theta}(t_1^{(j)}, t_2^{(j)}, \dots, t_n^{(j)})$ for each sample. The mean of the replicates is given as

SURVIVAL ESTIMATION USING THREE METHODS

$$\hat{\theta}_B = \frac{1}{m} \sum_{i=1}^m \hat{\theta}^{(j)}. \quad (3)$$

The mean squared error (MSE) is defined by $MSE(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right]$. If m random samples $t^{(1)}, t^{(2)}, \dots, t^{(m)}$ are generated from the distribution of T then estimate of the MSE of $\hat{\theta} = \hat{\theta}(t_1, t_2, \dots, t_n)$ is

$$MSE(\hat{\theta}_B) = \frac{1}{m-1} \sum_{j=1}^m (\hat{\theta}^{(j)} - \theta)^2 \quad (4)$$

where $\hat{\theta}^{(j)} = \hat{\theta}(t_1^{(j)}, t_2^{(j)}, \dots, t_n^{(j)})$.

Estimate of the standard error of the bootstrap estimate, $\hat{\theta}_B$ is given by

$$SE(\hat{\theta}_B) = \sqrt{\frac{1}{m-1} \sum_{j=1}^m (\hat{\theta}^{(j)} - \theta)^2} \quad (5)$$

100(1- α)% confidence interval for θ is given by

$$\hat{\theta}_B \pm Z_{\alpha/2} SE(\hat{\theta}_B) \quad (6)$$

The mean percentage error (MPE) is

$$MPE(\hat{\theta}_B) = \frac{\left[\sum_{j=1}^m \frac{|\hat{\theta}^{(j)} - \theta|}{\theta} \right]}{m}. \quad (7)$$

If the bootstrap estimator $\hat{\theta}_B$ is known from (3) then the estimate of survival function is given as

$$\hat{S}_B(t) = \exp\left(-\frac{t_i - t_0}{\hat{\theta}_B}\right) \quad (8)$$

Jackknife Estimation

The jackknife is a more orderly version of the bootstrap. As opposed to re-sampling randomly from the entire sample like the bootstrap does, the jackknife takes the entire sample except for 1 value, and then calculates the test statistic of interest. It repeats the process, each time leaving out a different value, and each time recalculating the test statistic. This method was introduced by Quenouille (1949) and further modification in Quenouille (1956). The theory is as follows:

Let $\hat{\theta}$ be an estimator of the parameter θ based on the complete sample of size n with g subgroups. Let $\hat{\theta}_{-i}$ be the corresponding estimator based on the sample at the i^{th} deletion. Define

$$\tilde{\theta}_i = g\hat{\theta} - (g-1)\hat{\theta}_{-i} \quad (i=1, 2, \dots, g) \quad (9)$$

The i^{th} deletion of the total could be one individual observation or several observation. The latter case is called group- or block-based jackknife if one replication or one block observations are deleted. In equation (1) estimation $\tilde{\theta}_i$ is called the i^{th} pseudo value and the estimator in equation (2) is the jackknife estimator for the parameter θ , where θ can be a variance component, covariance component, correlation coefficient, or any other parameter of interest.

$$\tilde{\theta} = \frac{1}{g} \sum_{j=1}^g \tilde{\theta}_j = g\hat{\theta} - (g-1) \frac{1}{g} \sum_{j=1}^g \hat{\theta}_{-j} \quad (10)$$

In equation (10), $\tilde{\theta}$ is called a pseudo jackknife estimate. A t-test can then be used to test significant deviation from a given parameter value, θ_0 with degrees of freedom $g-1$ (Miller, 1974a, b). The equation (9) can be rewritten as

$$\tilde{\theta}_i = g\hat{\theta} - (g-1)\hat{\theta}_{-i} = \hat{\theta} + (g-1)(\hat{\theta} - \hat{\theta}_{-i}) \quad (i=1, 2, \dots, g). \quad (11)$$

Thus, it is obvious that pseudo value $\tilde{\theta}_i$ in equation (11) is related to choices for g . When g is large, a slight difference between $\hat{\theta}$ and $\hat{\theta}_{-i}$ will cause unfavorable values. More importantly, it will potentially cause a large standard error for an estimate and thus decrease the power for the parameter being tested. If it is assumed that the estimate $\hat{\theta}_{-i}$ in equation (9) for the i^{th} deletion is unbiased,

SURVIVAL ESTIMATION USING THREE METHODS

then it is easy to prove that $\bar{\hat{\theta}}$ in equation (12) is unbiased too. It is often true if $\hat{\theta}$ is an unbiased estimate of θ , then $\hat{\theta}_{-i}$ will be unbiased after a few individuals in the original data are deleted.

$$\bar{\hat{\theta}}_{jack} = \frac{\sum_{i=1}^g \hat{\theta}_{-i}}{g}. \quad (12)$$

In equation (12), $\bar{\hat{\theta}}$ is called a non-pseudo jackknife estimate of the parameter θ . For each non-normally distributed variable, based on the Central Limit Theorem, $\bar{\hat{\theta}}$ is approximately normally distributed when g is large. Thus, an approximate z-test can be used when g is large or t-test can be used to test significant deviation from a given parameter value, θ_0 , with the degrees of freedom $g-1$. An estimate of the mean square error (MSE) of the jackknife estimate, $\bar{\hat{\theta}}_{jack}$ is given by

$$MSE(\bar{\hat{\theta}}_{jack}) = \frac{g}{g-1} \sum_{i=1}^g (\hat{\theta}_{-i} - \bar{\hat{\theta}})^2 \quad (13)$$

Estimate of the standard error of the jackknife estimate, $\bar{\hat{\theta}}_{jack}$ is given by

$$SE(\bar{\hat{\theta}}_{jack}) = \sqrt{\frac{g}{g-1} \sum_{i=1}^g (\hat{\theta}_{-i} - \bar{\hat{\theta}})^2} \quad (14)$$

100(1- α)% confidence interval for θ is given by

$$\bar{\hat{\theta}}_{jack} \pm t_{\alpha/2, g-1} SE(\bar{\hat{\theta}}_{jack})$$

The mean percentage error (MPE) is

$$MPE(\hat{\theta}_{jack}) = \frac{\left[\sum_{i=1}^g \frac{|\hat{\theta}_{-i} - \bar{\hat{\theta}}|}{\bar{\hat{\theta}}} \right]}{g}. \quad (15)$$

If the jackknife estimator $\hat{\theta}_{jack}$ is known from (12) then the estimate of survival function is given as

$$\hat{S}_{jack}(t) = \exp\left(-\frac{t_i - t_0}{\hat{\theta}_{jack}}\right). \quad (16)$$

K – Repeated Jackknife Estimation Method

The K – Repeated jackknife procedure is a re-sampling iterative scheme for mean square error (MSE) reduction. This involves jackknifing the observed data k -time, where k equals the sample size of the observed data. The procedure is conveniently applied when the sample size is small. The stopping rule for the repeated jackknife replications depends on the sample size of the original data. The procedure converges before or at k^{th} time, where the estimate from the jackknife replications is the same as estimator of the parameter θ based on the complete sample of size n . At the K^{th} time, the k^{th} – repeated jackknife estimate of bias is highly negligible.

The method involves the following steps from the usual jackknife procedure:

Step 1. Observe a random sample $T = (t_1, t_2, \dots, t_n)$

Step 2. Compute $\hat{\theta}(t)$ a function of the data which estimates the parameter θ of the model.

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n t_i \quad i = 1, 2, \dots, n \quad (17)$$

Step 3. For i up to n

- generate a jackknife sample $T_{-i} = (t_1, t_{i-1}, t_{i+1}, \dots, t_n)$ by leaving out the i^{th} observation
- calculate $\hat{\theta}_{-i}$ from each of the Jackknife sample T_{-i} by

$$\hat{\theta}_{-i} = \frac{1}{n-1} \sum_{i=1}^{n-1} T_{-i} \quad (18)$$

SURVIVAL ESTIMATION USING THREE METHODS

Step 4. Repeat step 3 using the estimates from $\hat{\theta}_{-i}$ to form pseudo samples. The new pseudo samples are used to generate another set of jackknife estimates; this is continued until the k^{th} time. This implies that the process is repeated k times, and at any given stage the preceding jackknife estimates are used as new samples in the next stage until the k^{th} time.

Step 5. At the k^{th} time the K -repeated Jackknife estimate is calculated as

$$\bar{\hat{\theta}}^K = \frac{1}{k} \sum_{i=1}^n \hat{\theta}_{i-1}^K \quad (19)$$

The K – repeated jackknife estimate of mean square error (MSE) is given by

$$MSE(\bar{\hat{\theta}}^K) = \frac{1}{k(k-1)} \sum_{i=1}^n \left(\hat{\theta}_{i-1}^K - \bar{\hat{\theta}}^K \right)^2 \quad (20)$$

The K – repeated jackknife estimate of standard error is given by

$$SE\left(\bar{\hat{\theta}}^K\right) = \sqrt{\frac{1}{k(k-1)} \sum_{i=1}^n \left(\hat{\theta}_{i-1}^K - \bar{\hat{\theta}}^K \right)^2} \quad (21)$$

An approximate $(1-\alpha)\%$ confidence interval for θ is given by

$$\bar{\hat{\theta}}^K \pm t_{\alpha/2, K-1} SE\left(\bar{\hat{\theta}}^K\right) \quad (22)$$

The mean percentage error (MPE) is

$$MPE\left(\bar{\hat{\theta}}^K\right) = \frac{\left[\sum_{i=1}^n \frac{\left| \hat{\theta}_{i-1}^K - \bar{\hat{\theta}}^K \right|}{\bar{\hat{\theta}}^K} \right]}{k}. \quad (23)$$

If the K – repeated jackknife estimator $\bar{\hat{\theta}}^K$ is known from (19), then the estimate of survival function is

$$\hat{S}^K(t) = \exp\left(-\frac{t_i - t_0}{\bar{\hat{\theta}}^K}\right). \quad (24)$$

The general iterative scheme is as follows: from a random sample $T = (t_1, t_2, \dots, t_n)$

$$\begin{aligned} 1. \quad \hat{\theta}_{(-1)}^1 &= \frac{1}{n-1} \sum_{i=1}^{n-1} T_{-i} \\ 2. \quad \hat{\theta}_{(-1)}^2 &= \frac{1}{n-1} \sum_{i=1}^{n-1} \hat{\theta}_{-i}^1 \\ 3. \quad \hat{\theta}_{(-1)}^3 &= \frac{1}{n-1} \sum_{i=1}^{n-1} \hat{\theta}_{-i}^2 \\ &\vdots \\ &\vdots \\ K. \quad \hat{\theta}_{(-1)}^K &= \frac{1}{n-1} \sum_{i=1}^{n-1} \hat{\theta}_{-i}^{K-1} \end{aligned}$$

Thus,

$$\bar{\hat{\theta}}^K = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_{i-1}^K$$

where $K = n$ (sample size) indicates the stopping rule. Other estimators such as variance, standard error and confidence interval can be estimated as in (20), (21), (22) and (23).

This study described three types of parameter estimation methods based on re-sampling technique: the bootstrap method, the jackknife method and the k -repeated jackknife method. However, the intention of this study is to use Monte Carlo simulated data to compare the three methods based on mean squared error (MSE) and mean percentage error (MPE), hence survival estimation.

Data and Analysis

Exploratory data analysis approach using simulated data generated by R-statistical program is adopted in this research work. This is to validate the statistical assumptions of an exponential distribution. In statistics, every statistical model has its own assumptions that have to be verified and met, to provide valid results. In the case of exponential distribution, the confidence interval for the mean life of an event requires two major assumptions: the time-to-occurrence of events of interest are independent, and the time for occurrence of event is exponentially distributed. These two statistical assumptions must be satisfied for the corresponding confidence interval to cover the true mean with the prescribed probability. The simulated data is based on random generation of values which satisfies both the assumption of independence and exponentially identical distribution. Some properties of the exponential distribution are as follows: the theoretical mean and standard deviation are equal. Hence, (1) the sample values of mean and standard deviation should be close. (2) Histogram should show that the distribution is right skewed (Median < Mean). (3) A plot of Cumulative-Failure vs. Cumulative-Time should be close to linear. (4) The regression slope of Cum-Failure vs. Cum-Time is close to the failure rate. (5) A plot of Cum-Rate vs. Cum-Failure should decrease/stabilize at the failure rate level. (6) Plots of the Exponential probability and its scores should also be close to linear. Some of these properties are explained by the exploratory data analysis displayed in Figures 1, i - xii.

Table 1. Generation Parameters

Sample 1 (Sample size = 10, $\lambda = 0.5$)	Sample 7 (Sample size = 20, $\lambda = 1.5$)
Sample 2 (Sample size = 10, $\lambda = 1.0$)	Sample 8 (Sample size = 20, $\lambda = 2.0$)
Sample 3 (Sample size = 10, $\lambda = 1.5$)	Sample 9 (Sample size = 30, $\lambda = 0.5$)
Sample 4 (Sample size = 10, $\lambda = 2.0$)	Sample 10 (Sample size = 30, $\lambda = 1.0$)
Sample 5 (Sample size = 20, $\lambda = 0.5$)	Sample 11 (Sample size = 30, $\lambda = 1.5$)
Sample 6 (Sample size = 20, $\lambda = 1.0$)	Sample 12 (Sample size = 30, $\lambda = 2.0$)

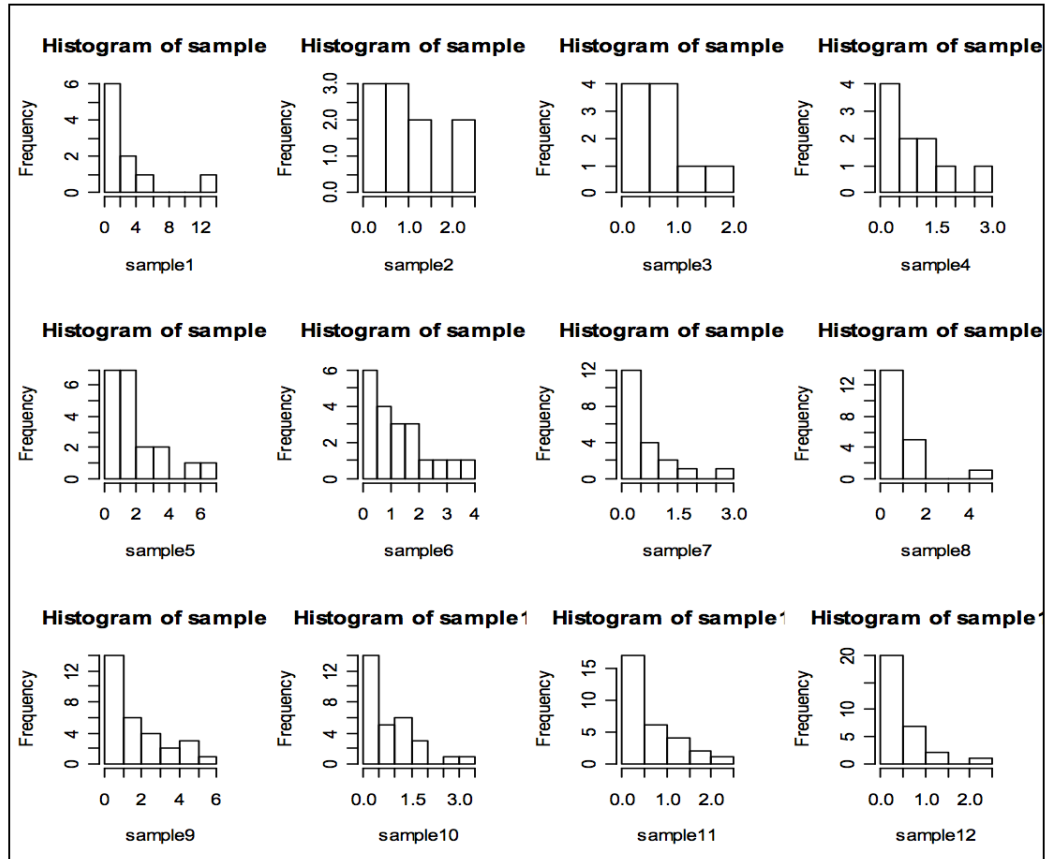


Figure 1. Histogram for each of the Randomly Generated Sample (i – xii)

SURVIVAL ESTIMATION USING THREE METHODS

Table 2. Descriptive Statistics of samples 1 to 12

Sample (i)	N	λ	Mean	Median	Std.Dev.
Sample 1	10	0.5	2.6006960	1.4973290	3.7961510
Sample 2	10	1.0	0.9211933	0.7875227	0.7792381
Sample 3	10	1.5	0.7320056	0.5990053	0.5961606
Sample 4	10	2.0	0.8510537	0.7067248	0.8785794
Sample 5	20	0.5	1.8608660	1.1843510	1.8059580
Sample 6	20	1.0	1.2379810	0.9922479	1.0967930
Sample 7	20	1.5	0.5363318	0.2470020	0.6540290
Sample 8	20	2.0	0.7281072	0.2824484	1.0112260
Sample 9	30	0.5	1.5960620	1.0194960	1.6278560
Sample 10	30	1.0	0.8559385	0.5422463	0.8296031
Sample 11	30	1.5	0.6353941	0.3864003	0.5880883
Sample 12	30	2.0	0.4639596	0.2700518	0.5150784

Results

The results of descriptive statistics show that as the sample sizes 10, 20 and 30 increase the mean and standard deviation are decreasing which satisfied one of the properties of that the theoretical mean and standard deviation are equal. The sample mean and standard deviation obtained are very close also as the value of λ increases the median values obtained get smaller. Figure 1 above shows that the observed distribution agrees with the exponential distribution property 1 and property 2 described in the data above. Figures 1, i–xii show right-skewness, which supported the attribute of an exponential distribution.

Table 3 shows the parameter estimation of the three methods. The results reveal that the estimation of the bootstrap approach is better than the other two methods that is the jackknifing and K repeated jackknifing. Table 4 is the result of the mean square error (MSE) of the analysis which is about the variance of the three methods. Results reveal that, as λ values increase, the results of jackknifing and K repeated jackknifing are better than the bootstrapped approach. Table 5 is the computation of the mean percentage error (MPE) the result shows that estimation of the bootstrap approach is better than the other two methods.

Table 3. Estimation Using the Three Methods Bootstrap, Jackknifing and K repeated jackknifing

	λ	$\hat{S}_B(t)$	$\hat{S}_{jack}(t)$	$\hat{S}^K(t)$
10	0.5	0.568858094	0.568879887	0.568879887
	1	0.476453626	0.476456925	0.476456925
	1.5	0.461343936	0.461328523	0.461328523
	2.0	0.529933691	0.529937819	0.529937819
20	0.5	0.491722891	0.491777729	0.491777729
	1	0.490229963	0.490240047	0.490240047
	1.5	0.544947075	0.544930402	0.544930402
	2.0	0.553586925	0.553580134	0.553580134
30	0.5	0.527441921	0.527445588	0.527445588
	1	0.491819455	0.491882638	0.491882638
	1.5	0.491085203	0.491099760	0.491099760
	2.0	0.528125037	0.528118624	0.528118624

Table 4. Estimation to the Bootstrap, Jackknifing and K repeated jackknifing using MSE methods

	λ	$\hat{S}_B(t)$	$\hat{S}_{jack}(t)$	$\hat{S}^K(t)$
10	0.5	0.004741437	0.004744439	0.004744439
	1	0.000554432	0.000554276	0.000554276
	1.5	0.001494291	0.001495483	0.001495483
	2.0	0.000896026	0.000896273	0.000896273
20	0.5	0.000068511	0.000067606	0.000067606
	1	0.000095454	0.000095257	0.000095257
	1.5	0.002020240	0.002018741	0.002018741
	2.0	0.002871559	0.002870831	0.002870831
30	0.5	0.000753059	0.000753260	0.000753260
	1	0.000066921	0.000065892	0.000065892
	1.5	0.000079474	0.000079214	0.000079214
	2.0	0.000791018	0.000790657	0.000790657

SURVIVAL ESTIMATION USING THREE METHODS

Table 5. Bootstrap, Jackknifing and K-repeated jackknifing using MPE methods

	λ	$\hat{S}_B(t)$	$\hat{S}_{jack}(t)$	$\hat{S}^K(t)$
10	0.5	0.137716188	0.137759774	0.137759774
	1	0.047092748	0.047086150	0.047086150
	1.5	0.077312128	0.077342954	0.077342954
	2.0	0.059867382	0.059875638	0.059875638
20	0.5	0.016554218	0.016444542	0.016444542
	1	0.019540074	0.019519906	0.019519906
	1.5	0.089894150	0.089860804	0.089860804
	2.0	0.107173850	0.107160268	0.107160268
30	0.5	0.054883842	0.054891176	0.054891176
	1	0.016361090	0.016234724	0.016234724
	1.5	0.017829594	0.017800480	0.017800480
	2.0	0.056250074	0.056237248	0.056237248

Table 6. Survival Estimation Using the Three Methods with Respect to MSE and MPE

BOOTSTRAP METHOD (1)					
Size	λ	$\hat{S}_B(t)$	MSE	MPE	REMARK
10	0.5	0.568858094	0.004741437	0.137716188	1
	1	0.476453626	0.000554432	0.047092748	2,3
	1.5	0.461343936	0.001494291	0.077312128	1
	2.0	0.529933691	0.000896026	0.059867382	1
20	0.5	0.491722891	0.000068511	0.016554218	2,3
	1	0.490229963	0.000095454	0.019540074	2,3
	1.5	0.544947075	0.002020240	0.089894150	2,3
	2.0	0.553586925	0.002871559	0.107173850	2,3
30	0.5	0.527441921	0.000753059	0.054883842	1
	1	0.491819455	0.000066921	0.016361090	2,3
	1.5	0.491085203	0.000079474	0.017829594	2,3
	2.0	0.528125037	0.000791018	0.056250074	2,3

Table 6, cont'd. Survival Estimation Using the Three Methods with Respect to MSE and MPE

JACKKNIFE METHOD (2)					
Size	λ	$\hat{S}_{jack}(t)$	MSE	MPE	REMARK
10	0.5	0.568858094	0.004741437	0.137716188	1
	1	0.476453626	0.000554432	0.047092748	2,3
	1.5	0.461343936	0.001494291	0.077312128	1
	2.0	0.529933691	0.000896026	0.059867382	1
20	0.5	0.491722891	0.000068511	0.016554218	2,3
	1	0.490229963	0.000095454	0.019540074	2,3
	1.5	0.544947075	0.002020240	0.089894150	2,3
	2.0	0.553586925	0.002871559	0.107173850	2,3
30	0.5	0.527441921	0.000753059	0.054883842	1
	1	0.491819455	0.000066921	0.016361090	2,3
	1.5	0.491085203	0.000079474	0.017829594	2,3
	2.0	0.528125037	0.000791018	0.056250074	2,3

K-REPEATED JACKKNIFE METHOD (3)					
Size	λ	$\hat{S}^K(t)$	MSE	MPE	REMARK
10	0.5	0.568858094	0.004741437	0.137716188	1
	1	0.476453626	0.000554432	0.047092748	2,3
	1.5	0.461343936	0.001494291	0.077312128	1
	2.0	0.529933691	0.000896026	0.059867382	1
20	0.5	0.491722891	0.000068511	0.016554218	2,3
	1	0.490229963	0.000095454	0.019540074	2,3
	1.5	0.544947075	0.002020240	0.089894150	2,3
	2.0	0.553586925	0.002871559	0.107173850	2,3
30	0.5	0.527441921	0.000753059	0.054883842	1
	1	0.491819455	0.000066921	0.016361090	2,3
	1.5	0.491085203	0.000079474	0.017829594	2,3
	2.0	0.528125037	0.000791018	0.056250074	2,3

Discussion

The comparison of parametric estimators of exponential distribution using Bootstrap, Jackknife, and K -Repeated Jackknife methods indicates that the estimates of the population parameter are very close which implies that the estimators are unbiased. A comparison of the mean square error (MSE) and mean percentage error (MPE) of the estimators shows that K -Repeated Jackknife method has a minimum variance unbiased estimator (MVUE); irrespective of the sample size whether it is small or large at any given values of lambda (λ). The three methods are used to estimate the survival function for exponential distribution and its mean square error (MSE) and mean percentage error (MPE). The results can be deduced that the performance of the two jackknife procedures over the bootstrap procedure is 66.67% to 33.33%. This result has been able to show the effect or influence of jackknife method, especially the k -repeated procedure on error reduction in estimating population parameter.

Conclusion

This study demonstrates that both methods of re-sampling technique are very efficient in estimating the population parameters and their mean square errors (MSE), as viewed by Efron (1998). These methods were used to find the best minimum variance unbiased estimator, using mean square error (MSE) and mean percentage error (MPE). The estimates of the two-parameter exponential distribution are used to estimate the survival probability. The attractiveness of jackknifing and bootstrapping is that they provide investigators with an important and unattainable type of information. Jackknifing and bootstrapping have their limitations and inherent assumptions as all statistical procedures do. The three methods are computationally intensive. However, these techniques represent an important step in refining the process of data analysis more especially the k -repeated procedure. Hence, it can be deduced that bootstrapping is a method for evaluating the variance of an estimator while jackknife is a method for reducing the bias of an estimator, and evaluating the variance of an estimator. This is clearly shown in the MSE results. The MSE value is reduced using the K -repeated jackknife method.

References

- Bradley, J. U. (1968). *Distribution-free statistical tests*. Englewood Cliff, NJ: Prentice-Hall.
- Collett, D. (1994). *Modeling survival data in medical research*. London: Chapman and Hall.
- Conover, W. J. (1980). *Practical nonparametric statistics*. New York: John Wiley and Sons.
- Cox, D. R., & Oakes, D. (1984). *Analysis of survival data*. New York: Chapman and Hall.
- Crowley, P. (1992). Resampling methods for computation-intensive data analysis in ecology and evolution. *Annual Review of Ecological Systems*, 23, 405-447.
- Efron, B., & Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, 37, 36-48.
- Efron, B., & Tibshirani, R. (1998). *An introduction to the bootstrap*. Boca Raton, FL: Chapman & Hall/CRC.
- Fernández, A. J. (2000). Estimation and Hypothesis Testing for Exponential Lifetime Models with Double Censoring and Prior Information. *Journal of Economic and Social Research*, 2(2), 1-17.
- Good, P. L. (2005). *Resampling methods* (3rd ed.). Berlin: Birkhauser.
- Lawless, J. F. (1982). *Statistical models and methods for lifetime data*. New York: John Wiley and Sons.
- Miller, R. G. (1974a). An unbalanced jackknife. *Annals of Statistics*, 2, 880-891.
- Miller, R. G. (1974b). The jackknife-a review. *Biometrika*, 61, 1-15.
- Quenouille, M. H. (1949). Approximate tests of correlation in time-series. *Journal of the Royal Statistical Society, Series B*, 11, 68-84.
- Quenouille, M. H. (1956). Notes on bias in estimation. *Biometrika*, 43, 353-360.
- Rizzo, M. L. (2008). *Statistical computing with R*. Boca Raton, FL: Chapman and Hall/CRC.
- Suzuki, K. (1985a). Nonparametric estimation of lifetime distributions from a record of failures and follow-ups. *Journal of the American Statistical Association*, 80, 68-72.

SURVIVAL ESTIMATION USING THREE METHODS

Suzuki, K. (1985b). Estimation of lifetime parameters from incomplete field data. *Technometrics*, 27, 263-271.

Appendix

Simulated Data

Table A1. Sample Size = 10

S/N	Sample 1($\lambda= 0.5$)	Sample 2($\lambda= 1.0$)	Sample 3($\lambda= 1.5$)	Sample 4($\lambda= 2.0$)
1	0.38672500	2.27994293	0.20192160	0.01058359
2	0.11665480	0.71411938	0.48301090	2.51951031
3	0.12661390	0.75298901	0.12473380	1.94495769
4	4.42286640	2.19289233	1.99662440	0.67531853
5	2.00704600	0.04442099	0.67780240	1.05647689
6	2.34689070	0.35021824	0.75981590	0.01373747
7	0.86211020	1.01674933	0.87373650	0.08979746
8	1.03118540	0.82205630	0.21055390	0.73813115
9	12.74339800	0.02251542	1.47164790	0.07333057
10	1.96347260	1.01602885	0.52020820	1.38869338

Table A2. Sample Size = 20

S/N	Sample 5($\lambda= 0.5$)	Sample 6($\lambda= 1.0$)	Sample 7($\lambda= 1.5$)	Sample 8($\lambda= 2.0$)
1	2.62129072	0.13858938	0.61025422	0.20816883
2	3.66009415	0.23677409	1.39562321	0.10508682
3	0.28671927	0.65882621	1.64585069	0.16343073
4	0.53614812	2.03913598	1.11779841	4.47429656
5	1.67262059	3.07310425	0.64918675	0.31910786
6	0.08521250	0.85047816	0.05778008	0.21707275
7	0.84341716	0.93587816	0.04898416	0.04360368
8	1.87988871	3.97283307	0.12098188	0.11505109
9	3.13213741	1.91406287	0.55509004	0.24578897
10	6.82508233	1.04861761	0.19238185	1.05875641
11	0.02900937	0.03350444	0.26953211	0.00355591
12	1.74900498	1.30175357	0.01133644	0.05922794
13	1.24431388	1.52538027	0.28231025	0.20815442
14	1.12438751	0.06504098	2.50193518	1.48214568
15	1.02901637	1.90192968	0.52524965	0.89694421
16	2.19818977	0.71117769	0.20890800	1.53327300
17	5.75705560	2.66119925	0.12742349	1.11055824
18	0.67605113	1.38839741	0.09933566	0.81368075
19	0.79778248	0.30239189	0.08220231	0.46919296
20	1.06989471	0.00054309	0.22447181	1.03504680

SURVIVAL ESTIMATION USING THREE METHODS

Table A3. Sample Size = 30

S/N	Sample 9($\lambda= 0.5$)	Sample 10($\lambda= 1.0$)	Sample 11($\lambda= 1.5$)	Sample 12($\lambda= 2.0$)
1	1.09871091	0.17085314	0.24591320	0.03006484
2	1.30597223	0.31266839	0.78554402	0.01316790
3	0.43742079	0.21960578	0.26193004	0.08580695
4	0.57034193	1.09603105	0.07651001	0.06684230
5	1.02475744	0.54796524	0.29620540	0.75349995
6	1.53172531	0.13561858	0.58560088	0.56383185
7	0.57173448	0.46805576	0.56401687	0.10014385
8	3.18835121	0.58675535	0.41054238	0.39468972
9	1.01423546	0.41116558	1.42498655	0.26076246
10	0.05104055	3.42353290	0.03212381	0.38563026
11	2.99399940	1.09817464	0.19768464	0.96558979
12	2.95112802	0.35962685	0.24295821	0.27934118
13	4.74244122	1.98862082	0.16056365	0.01547041
14	0.04628853	1.11964839	0.24140637	0.34343548
15	5.35809191	0.48539163	0.06956623	0.80092480
16	4.26185504	1.09365222	1.49188159	1.49780414
17	0.04367701	0.73949713	0.22345808	0.90349970
18	0.05851474	0.34345758	1.56937093	0.18701462
19	0.46455153	0.29243694	0.59195204	0.00288255
20	0.09371455	0.25904558	0.42176436	0.74330912
21	2.96970220	1.79180600	1.11983745	0.25455636
22	0.54133977	0.01809066	0.28416911	0.81341579
23	1.29204462	2.94010580	0.65638599	0.10230827
24	0.02923826	0.79327459	0.12515747	2.24807427
25	2.28724168	1.13577062	2.26479793	0.08718112
26	3.62406597	0.05051664	1.73361763	1.15865739
27	0.55831489	0.53652729	0.36225826	0.02484439
28	4.24494133	1.08361205	1.41035274	0.48756216
29	0.13192788	0.25488727	0.93471466	0.25596331
30	0.39449119	1.92176066	0.27655395	0.09251206

Life Testing Analysis of Failure Censored Generalized Exponentiated Data

Anwar Hassan

King Saud University
Riyadh, Saudi Arabia

Mehraj Ahmad

Directorate of Economics and Statistics
Jammu & Kashmir, India

A generalized exponential distribution is considered for analyzing lifetime data; such statistical models are applicable when the observations are available in an ordered manner. This study examines failure censored data, which consist of testing n items and terminating the experiment when a pre-assigned number of items, for example r ($< n$), have failed. Due to scale and shape parameters, both have flexibility for analyzing different types of lifetime data. This distribution has increasing, decreasing and a constant hazard rate depending on the shape parameter. This study provides maximum likelihood estimation and uniformly minimum variance unbiased techniques for the estimation of reliability of a component. Numerical computation was conducted on a data set and a comparison of the performance of two different techniques is presented.

Keywords: Generalized exponential distribution, lifetime data, censored data, uniformly minimum variance unbiased estimation

Introduction

Usually observations made on a random variable do not become available in an ordered manner. If n items are taken from a machine and measured for some characteristics such as diameter, it would be an anomaly – as well as a cause for concern – if the first item taken had the smallest diameter; the second item, the second smallest diameter, etc. However, there exist numerous practical situations, for example, life testing fatigue and other kinds of destructive test situations, where the data become available in this way. If n radio tubes are put through a life test, for example, then the weakest will fail first in time, the second weakest one fails next, etc. Based on this pattern, it seems clear that observations will naturally occur in an ordered manner in life test situations, regardless of whether the test is the life of electric bulbs, life of radio tubes, life of ball bearings, life of various kinds of

Dr. Hassan is in the Department of Statistics and O. R., Faculty of Science. Email him at: anwar.hassan2007@gmail.com. Mehraj serves in the Government of Jammu and Kashmir, India.

physical equipment or length of life after some treatment performed on animals or human beings. There are other situations – for example, testing the current needed to blow out a fuse, the voltage needed to break down a condenser, the force needed to rupture some physical material, etc. – where observations become available in order if the test is arranged in such a way that every item in the sample is subjected to the same stimulus (current, voltage, stress, dosage, etc.), so that the first weakest item fails, then the second weakest item fails, and so on.

Put in general terms, if n items drawn at random from some generalized exponential population are tested, and the data become available in such a way that the smallest observation comes first, the second smallest second, and so on until finally the largest observation is last, then it is possible to discontinue experimentation after observing the first r failures in a life test. The two principal advantages associated with the possibility of stopping before all n observations are made stem from the observations occurring in an ordered manner and the ability to reach a decision in a shorter time or with fewer observations than if utilizing a procedure that involves observing what happens to all items being tested. Thus, this study is devoted to failure censored data, which consists of putting n items on test and terminating the experiment when a pre-assigned number of items, for example r ($< n$), have failed. The data obtained from such experimentation is almost mandatory in dealing with high cost sophisticated items such as televisions.

The Generalized Exponential Distribution (GED), which more accurately represents time to failure, is used instead of the more commonly used exponential distribution. Although incorporation of the GED in life testing modeling adds to the complexity of modeling and estimation, it fits life data more accurately than the exponential distribution due to its flexibility.

The two parameter GED was proposed and studied extensively by Gupta and Kundu (1999, 2001a, 2001b, 2002), Raqab (2002), Raqab and Ahsanullah (2001) and Zheng (2002) and the two parameter GED distribution has: density function

$$f(x, \alpha, \lambda) = \alpha \lambda (1 - e^{-\lambda x})^{\alpha-1} e^{-\lambda x}, x > 0, \alpha > 0, \lambda > 0, \quad (1)$$

cumulative distribution function (cdf)

$$F(x, \alpha, \lambda) = (1 - e^{-\lambda x})^{\alpha}, x > 0, \alpha > 0, \lambda > 0, \quad (2)$$

survival function

$$S(x, \alpha, \lambda) = 1 - (1 - e^{-\lambda x})^\alpha, x > 0, \alpha > 0, \lambda > 0, \quad (3)$$

and hazard function

$$h(x, \alpha, \lambda) = \frac{\alpha \lambda (1 - e^{-\lambda x})^{\alpha-1} e^{-\lambda x}}{1 - (1 - e^{-\lambda x})^\alpha}, x > 0, \alpha > 0, \lambda > 0. \quad (4)$$

Here $\alpha > 0$ and $\lambda > 0$ are the shape and scale parameters respectively. For different values of the shape parameter, the density function can take different shapes. Hereafter, the GED with shape parameter α and scale parameter λ will be denoted by $GE(\alpha, \lambda)$. This article focuses on the maximum likelihood estimate and the minimum variance unbiased estimate of the shape parameter when the scale parameter is known.

Estimation Based on MLE

Maximum Likelihood Estimation

Suppose n items are subjected to test without replacement and the test is terminated after r items have failed. If the failure censored data consist of the lifetimes of the r items that failed ($X_{(1)} < X_{(2)} < \dots < X_{(r)}$) and the fact that $(n - r)$ items have survived beyond $X_{(r)}$. The likelihood of the ordered sample failure times is given below if the failure times are generalized exponentially distributed with pdf (1).

For given ordered failures times when it is desired to estimate α when λ is known:

$$L(\alpha, \lambda | x) = \binom{n}{r} r! \prod_{i=1}^r \alpha \lambda (1 - e^{-\lambda x_i})^{\alpha-1} e^{-\lambda x_i} \prod_{i=r+1}^n (1 - e^{-\lambda x_r})^\alpha$$

$$L(\alpha, \lambda | x) = \binom{n}{r} r! \alpha^r \lambda^r \prod_{i=1}^r (1 - e^{-\lambda x_i})^{\alpha-1} e^{-\lambda \sum_{i=1}^r x_i} (1 - e^{-\lambda x_r})^{\alpha(n-r)}.$$

The log likelihood function of the observed sample is

$$L(\alpha) = \ln(C) + r \ln(\alpha) + r \ln(\lambda) - \lambda \sum_{i=1}^r x_i + (\alpha - 1) \sum_{i=1}^r \ln(1 - e^{-\lambda x_i}) \alpha (n - r) \ln(1 - e^{-\lambda x_r}) \quad (5)$$

where $C = \binom{n}{r} r!$

The MLE of α , for example, $\hat{\alpha}$ for known λ is

$$\begin{aligned} \hat{\alpha} &= - \frac{r}{\sum_{i=1}^r \ln(1 - e^{-\lambda x_i}) + (n - r) \ln(1 - e^{-\lambda x_r})} \\ &= \frac{r}{\sum_{i=1}^r T_i + (n - r) T_r} = \frac{r}{T' + (n - r) T''} \end{aligned}$$

where $T_i = \ln(1 - e^{-\lambda x_i})^{-1}$ and $T_r = \ln(1 - e^{-\lambda x_r})^{-1}$
and $T' = \sum_{i=1}^r \ln(1 - e^{-\lambda x_i})^{-1}$, $T'' = (1 - e^{-\lambda x_r})^{-1}$.

Unbiasedness of $\hat{\alpha}$

If the n items are tested and observation continues until r units have failed then $(T_{(1)}, T_{(2)}, \dots, T_{(r)})$ are the transferred failure time from exponential population with mean life α . Because X_1, X_2, \dots, X_r are independently and identically distributed (*iid*) $GED(\alpha, \lambda)$, then $T_{(i)}$, the transformed ordered failures, are *iid* as $\text{Expo}(\alpha)$. In this plan the number of items exposed at any time is n , the joint distribution of $T_{(1)}, T_{(2)}, \dots, T_{(r)}$, that is, the number of failed items out of n items tested is given by

$$g(t_{(1)}, t_{(2)}, \dots, t_{(r)} | \alpha) = \binom{n}{r} r! \alpha^r e^{-\sum_{i=1}^r t_{(i)}}; 0 < T_{(1)} < T_{(2)} < \dots < T_{(r)} < \infty$$

Using transformation $Z_i = (n - i + 1)(T_{(i)} - T_{(i-1)})$, $i = 1, 2, 3, \dots, r$, with $T_{(0)} = 0$, then

$$\sum_{i=1}^r Z_i = \sum_{i=1}^r T_{(i)} + (n-r)T_{(r)} \quad (6)$$

$$\frac{\partial(Z_1, Z_2, \dots, Z_r)}{\partial(T_{(1)}, T_{(2)}, \dots, T_{(r)})} = \binom{n}{r} r! \text{ and } |J| = \frac{1}{\binom{n}{r} r!}$$

This results in the joint distribution of Z_1, Z_2, \dots, Z_r as $g(Z_1, Z_2, \dots, Z_r | \alpha) = \alpha^r e^{-\sum_{i=1}^r Z_i}$, thus Z_1, Z_2, \dots, Z_r are iid as $g(z | \alpha) = \alpha e^{-\alpha z}$; $z, \alpha \geq 0$

$$\hat{\alpha} = \frac{r}{\sum_{i=1}^r Z_i}$$

$$\frac{\hat{\alpha}}{r\alpha} = \frac{1}{\alpha \sum_{i=1}^r Z_i} \text{ or } \frac{r\alpha}{\hat{\alpha}} = \alpha \sum_{i=1}^r Z_i \text{ follows Gamma}(r)$$

and $Y = \frac{\hat{\alpha}}{r\alpha} = \frac{1}{\alpha \sum_{i=1}^r Z_i}$ follows inverted gamma density of Raiffa and Schlaifer (1961) as

$$f(y) = \frac{1}{\Gamma r} e^{-\frac{1}{y}} \left(\frac{1}{y} \right)^{r+1}; y \geq 0, r \geq 0, \quad (7)$$

$$\text{and the pdf of } \hat{\alpha} \text{ is } f(\hat{\alpha} | \alpha) = \frac{1}{\alpha \Gamma(r+1)} e^{-\frac{r\alpha}{\hat{\alpha}}} \left(\frac{r\alpha}{\hat{\alpha}} \right)^{r+1}; \hat{\alpha} \geq 0, r \geq 0. \quad (8)$$

Moments of $\hat{\alpha}$

It is necessary to extract the first two moments of $\hat{\alpha}$, to find in general the k^{th} moment of $\hat{\alpha}$ as

$$\mu'_k = E(Y^k) = \int_0^\infty y^k \frac{1}{\Gamma r} e^{-\frac{1}{y}} \left(\frac{1}{y} \right)^{r+1} dy = \frac{\Gamma(r-k)}{\Gamma r}$$

$$\mu'_1 = \frac{1}{(r-1)}, \mu'_2 = \frac{1}{(r-1)(r-2)}, \text{ and } \mu_2 = \frac{1}{(r-1)^2(r-2)}$$

$$\mu'_1 = E(Y) = E\left(\frac{\hat{\alpha}}{r\alpha}\right) = \frac{1}{(r-1)} \text{ results in } E(\hat{\alpha}) = \frac{r\alpha}{r-1}$$

$$\mu_2 = V(Y) = V\left(\frac{\hat{\alpha}}{r\alpha}\right) = \frac{1}{(r-1)^2(r-2)} \text{ results in } V(\hat{\alpha}) = \frac{r^2\alpha^2}{(r-1)^2(r-2)}$$

Thus $E(\hat{\alpha}) \neq \alpha$, which clearly shows that the MLE of α is not an unbiased estimate of α , but instead it is asymptotically unbiased estimate of α .

Sufficiency of $\hat{\alpha}$

$$L(x_1, x_2, \dots, x_r | \alpha, \lambda) = \frac{n!}{(n-r)!} \alpha^r \lambda^r \prod_{i=1}^r (1 - e^{-\lambda x_i})^{\alpha-1} \prod_{i=1}^r e^{-\lambda x_i} (1 - e^{-\lambda x_r})^{\alpha(n-r)}$$

Using the transformation as in [Lemma 1](#) (see [Appendix A](#)) results in,

$$L(x_1, x_2, \dots, x_r | \alpha, \lambda) = \frac{n!}{(n-r)!} \alpha^r \lambda^r e^{-\alpha[\sum_{i=1}^r t_i + (n-r)T_r]} e^{\sum_{i=1}^r t_i} \prod_{i=1}^r (1 - e^{-t_i})$$

Using the transformation as in [\(6\)](#) results in,

$$L(x_1, x_2, \dots, x_r | \alpha, \lambda) = \alpha^r \lambda^r e^{-\alpha \sum_{i=1}^r z_i} \frac{n!}{(n-r)!} e^{\sum_{i=1}^r t_i} \prod_{i=1}^r (1 - e^{-t_i})$$

Using $\hat{\alpha} = \frac{r}{\sum_{i=1}^r T_i} = \frac{r}{\sum_{i=1}^r Z_i}$ results in,

$$\frac{L(x_1, x_2, \dots, x_r | \alpha, \lambda)}{f(\hat{\alpha} | \alpha)} = \frac{n!}{(n-r)!} \frac{\Gamma r}{r^r} \lambda^r \prod_{i=1}^r (1 - e^{-t_i}) e^{\sum_{i=1}^r t_i} (\hat{\alpha})^{(r+1)}$$

which is independent of the unknown parameter α , thus $\hat{\alpha}$ is a sufficient estimator for α .

MLE of Reliability

Because the MLE of α i.e. $\hat{\alpha} = \frac{r}{\sum_{i=1}^r T_i + (n-r)T_r} = \frac{r}{\sum_{i=1}^r Z_i}$ has been calculated using a property of MLE, that function of an MLE is also an MLE, thus the MLE of reliability of GED is denoted by $\hat{R}(t)$ and is given as $\hat{R}(t) = 1 - (1 - e^{-\lambda t})^{\hat{\alpha}}$.

Expectation of Reliability and its Standard Error

To evaluate the expectation of reliability and its standard error, results from Watson (1952) viz $\int_0^\infty x^{-r} e^{-\left(ax + \frac{b}{x}\right)} dx = 2 \left(\frac{a}{b}\right)^{\frac{r-1}{2}} K_{r-1}(2\sqrt{ab})$ are used where $K_r(\cdot)$ is the modified Bessels function of the second kind of order r .

$$E(\hat{R}(t)) = 1 - \frac{2}{\Gamma r} (\ln A_o^{-r\alpha})^{\frac{r}{2}} K_{-r}(2\sqrt{\ln A_o^{-r\alpha}})$$

and

$$E(\hat{R}(t))^2 = 2E(\hat{R}(t)) + \frac{2}{\Gamma r} (\ln A_o^{-2r\alpha})^{\frac{r}{2}} K_{-r}(2\sqrt{\ln A_o^{-2r\alpha}}) - 1.$$

Estimation Based on Minimum Variance Unbiased Estimate

Minimum Variance Unbiased Estimate

The Minimum Variance Unbiased Estimate (MVUE) approach is now considered. Note that $\hat{\alpha}$ is biased, but the bias can be easily corrected as

$$\tilde{\alpha} = \frac{r-1}{r} \hat{\alpha} = \frac{r-1}{r} \frac{r}{\sum_{i=1}^r Z_i} = \frac{r-1}{\sum_{i=1}^r Z_i}$$

this implies, $\frac{\tilde{\alpha}}{(r-1)\alpha} = \frac{1}{\alpha \sum_{i=1}^r Z_i}$

Recall the proven result that $\alpha \sum_{i=1}^r Z_i$ follows Gamma(r) and $Y = \frac{\tilde{\alpha}}{(r-1)\alpha} = \frac{1}{\alpha \sum_{i=1}^r Z_i}$ follows the inverted gamma density of Raiffa and Schlaifer (1961) as

$$f(y) = \frac{1}{\Gamma r} e^{-\frac{1}{y}} \left(\frac{1}{y} \right)^{r+1}; y \geq 0, r \geq 0;$$

the pdf of $\tilde{\alpha}$ is then:

$$f(\tilde{\alpha} | \alpha) = \frac{1}{\alpha(r-1)\Gamma r} e^{-\frac{(r-1)\alpha}{\tilde{\alpha}}} \left(\frac{(r-1)\alpha}{\tilde{\alpha}} \right)^{r+1}; \tilde{\alpha} \geq 0, r \geq 0$$

$$\mu'_k = E(Y^k) = \frac{\Gamma(r-k)}{\Gamma r}$$

$$\mu'_1 = \frac{1}{(r-1)}, \mu'_2 = \frac{1}{(r-1)(r-2)} \text{ and } \mu_2 = \frac{1}{(r-1)^2(r-2)}$$

$$\mu'_1 = E(Y) = E\left(\frac{\tilde{\alpha}}{(r-1)\alpha}\right) = \frac{1}{(r-1)} \text{ results in } E(\tilde{\alpha}) = \alpha$$

$$\mu_2 = V(Y) = V\left(\frac{\tilde{\alpha}}{(r-1)\alpha}\right) = \frac{1}{(r-1)^2(r-2)} \text{ results in } V(\tilde{\alpha}) = \frac{\alpha^2}{(r-2)}$$

Clearly $V(\hat{\alpha}) \geq V(\tilde{\alpha})$. However, equality holds for $r = \frac{1}{2}$ which is not an integer, thus it implies that this inequality never holds for integral value of n .

Sufficiency of $\tilde{\alpha}$

$$L(x_1, x_2, \dots, x_r | \alpha, \lambda) = \frac{n!}{(n-r)!} \alpha^r \lambda^r \prod_{i=1}^r (1 - e^{-\lambda x_i})^{\alpha-1} \prod_{i=1}^r e^{-\lambda x_i} (1 - e^{-\lambda x_r})^{\alpha(n-r)}$$

Using the transformation in Lemma 1 results in

$$L(x_1, x_2, \dots, x_r | \alpha, \lambda) = \frac{n!}{(n-r)!} \alpha^r \lambda^r e^{-\alpha [\sum_{i=1}^r t_i + (n-r)T_r]} e^{\sum_{i=1}^r t_i} \prod_{i=1}^r (1 - e^{-t_i})$$

Using the transformation in (6) results in

$$L(x_1, x_2, \dots, x_r | \alpha, \lambda) = \alpha^r \lambda^r e^{-\alpha \sum_{i=1}^r z_i} \frac{n!}{(n-r)!} e^{\sum_{i=1}^r t_i} \prod_{i=1}^r (1 - e^{-t_i})$$

Using $\frac{\tilde{\alpha}}{(r-1)} = \frac{1}{\sum_{i=1}^r z_i}$ results in

$$\frac{L(x_1, x_2, \dots, x_r | \alpha, \lambda)}{f(\tilde{\alpha} | \alpha)} = \frac{n!}{(n-r)!} \frac{\Gamma r}{(r-1)^r} \lambda^r \prod_{i=1}^r (1 - e^{-t_i}) e^{\sum_{i=1}^r t_i} (\tilde{\alpha})^{(r+1)}$$

which is independent of the unknown parameter α and, thus, $\tilde{\alpha}$ is a sufficient estimator for α .

Completeness

A family of density functions $f(X, \alpha), \alpha \in H$ (Parametric Space) is called complete if $E(u(x)) = 0$ for all $\alpha \in H$ implies $u(x) = 0$ with probability 1, for all $\alpha \in H$.

That is, there are no two different functions of X which have the same expected value for all $\alpha \in H$. Thus, for example, if a sufficient statistic is complete, there will be only one unbiased estimator of α which is a function of the sufficient statistic.

$$\begin{aligned} E_{\alpha}(\psi(\tilde{\alpha})) &= \int_0^{\infty} \psi(\tilde{\alpha}) \frac{1}{\alpha(r-1)\Gamma r} e^{\frac{-(r-1)\alpha}{\tilde{\alpha}}} \left(\frac{(r-1)\alpha}{\tilde{\alpha}} \right)^{r+1} d\tilde{\alpha} \\ &= \int_0^{\infty} \psi^*(\tilde{\alpha}) e^{\frac{-(r-1)\alpha}{\tilde{\alpha}}} d\tilde{\alpha} \end{aligned}$$

where $\psi^*(\tilde{\alpha})$ includes all other terms.

Now using Laplas transformation that $\int_0^{\infty} h(x) e^{-\theta x} dx = 0 \Rightarrow h(x) = 0$

$$E_{\alpha}(\psi(\tilde{\alpha})) = 0 \Rightarrow \psi^*(\tilde{\alpha}) = 0 \Rightarrow \psi(\tilde{\alpha}) = 0; \alpha \geq 0, r > 0$$

Thus $\tilde{\alpha}$ is also a complete estimate of α because $E(\tilde{\alpha}) = \alpha$, it follows that $\tilde{\alpha}$ is a uniformly minimum variance unbiased estimate (UMVUE) of α .

UMVUE of Reliability

Previously $\tilde{\alpha} = \frac{r-1}{\sum_{i=1}^r Z_i}$, which is an UMVUE was estimated, the UMVUE of reliability $\tilde{R}(t)$ is derived next. The general method of finding the UMVUE is to search for any unbiased statistics $T(x_1, x_2, \dots, x_n)$ and a complete and sufficient statistic if one exists. Consider a function $T(x_1, x_2, \dots, x_n)$ such that $T(x_1) = 1$ if $x_1 \geq t$ and $= 0$ otherwise. Thus, T is a function of x_1 alone, denoted by $T(x_1)$.

$$E\{T(x_1)\} = 1.P(X_1 \geq t) + 0.P(X_1 < t) = P(X_1 \geq t) = \tilde{R}(t|\alpha)$$

For its derivation it is necessary to derive the conditional distribution of $T_{(1)}$ given $\tilde{\alpha}$ and split the random sample $(T_{(1)}, T_{(2)}, \dots, T_{(r)})$ into two independent components $T = T_{(1)}$ of sample size one and $(T_{(2)}, T_{(3)}, \dots, T_{(r)})$ of sample size $(r-1)$. Because $T_{(i)}$ are *iid* as exponential with parameter α as proved in [Lemma 1](#), then

$$f(t_{(i)}|\alpha) = \alpha e^{-\alpha t_{(i)}}, \alpha \geq 0; t_{(i)} \geq 0$$

and define $\bar{Y} = \frac{1}{r-1} \sum_{i=2}^r Z_{(i)}$

$$(r-1)\bar{Y} + T_{(1)} = \sum_{i=1}^r T_{(i)} = r\bar{T}$$

This yields $(r-1)\bar{Y}$ follows $\text{Gamma}(\alpha, r-1)$ because $\sum_{i=1}^r T_{(i)} = \sum_{i=1}^r Z_i \sim \text{Gamma}(\alpha, r)$ and assumes $(r-1)\bar{Y} = S \sim \text{Gamma}(\alpha, r-1)$

$$f(\bar{y}) = \frac{\alpha^{r-1}}{\Gamma(r-1)} e^{-\alpha(r-1)\bar{y}} ((r-1)\bar{y})^{r-2} (r-1)$$

$$f(\bar{y}) = \frac{\alpha^{r-1} (r-1)^{(r-1)}}{\Gamma(r-1)} e^{-\alpha(r-1)\bar{y}} (\bar{y})^{r-2}$$

and then the joint distribution of $T_{(1)}$ and \bar{y} is given by

$$\begin{aligned} f(t_{(1)}, \bar{y} | \alpha) &= \alpha e^{-\alpha t_{(1)}} \frac{\alpha^{r-1} (r-1)^{(r-1)}}{\Gamma(r-1)} e^{-\alpha(r-1)\bar{y}} (\bar{y})^{r-2} \\ &= \frac{\alpha^n (r-1)^{(r-1)}}{\Gamma(r-1)} e^{-\alpha\{(r-1)\bar{y} + t_{(1)}\}} (\bar{y})^{r-2} \end{aligned}$$

Because $t_{(1)}$ and \bar{y} are independently distributed, the joint distribution of \bar{T} and $T_{(1)}$ can be obtained by using the transformation

$$(r-1)\bar{Y} + T_{(1)} = \sum_{i=1}^r T_{(i)} = r\bar{T}, \text{ resulting in } |J| = \frac{d\bar{y}}{d\bar{t}} = \frac{r}{r-1}$$

Therefore, $f(t_{(1)}, \bar{t} | \alpha) = \frac{\alpha^r r (r-1)^{(r-2)}}{\Gamma(r-1)} e^{-\alpha r \bar{t}} \left[\frac{r\bar{t}}{r-1} - \frac{t_{(1)}}{r-1} \right]^{r-2}; 0 < t_{(1)} < r\bar{t}$

$$\text{and } f(\bar{t} | \alpha) = \frac{(r\alpha)^r}{\Gamma r} (\bar{t})^{r-1} e^{-r\alpha\bar{t}}$$

The conditional distribution of $t_{(1)}$ is then obtained as $f(t_{(1)} | \bar{t}, \alpha) = \frac{f(t_{(1)}, \bar{t} | \alpha)}{f(\bar{t} | \alpha)}$

$$f(t_{(1)} | \bar{t}, \alpha) = \frac{r-1}{r\bar{t}} \left[1 - \frac{t_{(1)}}{r\bar{t}} \right]^{r-2} ; 0 < t_{(1)} < r\bar{t}$$

Using $\tilde{\alpha} = \frac{r-1}{\sum_{i=1}^r T_i}$ results in $\bar{t} = \frac{r-1}{nr}$ and using this value in the above pdf of $t_{(1)}$

$$f(t_{(1)} | \tilde{\alpha}) = \tilde{\alpha} \left[1 - \frac{t_{(1)}\tilde{\alpha}}{r-1} \right]^{r-2} ; 0 < t_{(1)} < \frac{r-1}{\tilde{\alpha}}$$

Thus UMVUE of reliability $\tilde{R}(t)$ is obtained as

$$\begin{aligned} \tilde{R}(t) &= P(X_1 \geq t_o | \tilde{\alpha}) = P\{-\ln(1 - e^{-\lambda x_1}) < -\ln(1 - e^{-\lambda t_o}) | \tilde{\alpha}\} \\ &= P\{T_1 < -\ln(1 - e^{-\lambda t_o}) | \tilde{\alpha}\} = \int_0^{-\ln(1 - e^{-\lambda t_o})} \tilde{\alpha} \left[1 - \frac{t_{(1)}\tilde{\alpha}}{r-1} \right]^{r-2} dt_{(1)} \end{aligned}$$

$$\tilde{R}(t) = 1 - \left\{ 1 + \frac{\tilde{\alpha} \ln(1 - e^{-\lambda t_o})}{r-1} \right\}^{r-1}$$

Expectation of MVUE Reliability and Its Standard Error

$$E(\tilde{R}(t)) = 1 - \sum_{j=0}^{r-1} \frac{(\alpha A)^j}{j!}$$

$$E(\tilde{R}(t))^2 = 2E(\tilde{R}(t)) + \frac{1}{\Gamma r} \sum_{j=0}^{2r-2} \left(\frac{2r-2}{j} \right) \Gamma(r-j) (\alpha A)^j - 1$$

Data Analysis

Sixty items were tested and the test was terminated after the first 10 items failed. The failure times (in months) were recorded as 0.12, 0.21, 0.39, 0.52, 0.68, 0.72, 0.87, 0.99, 1.14, 1.27. Assume that failure times are distributed as generalized exponentially distributed.

The mean value of failure times is 0.69 months. The parameter α and reliability was estimated using the MLE and MVUE for various known values of λ and the behavior of two different estimations on the estimation of reliability and parameter estimation was studied; results are shown in [Tables 1](#) and [2](#) (see [Appendix A](#)).

Acknowledgements

The project was supported by the Research Center, College of Science, King Saud University, Saudi Arabia.

References

- Gupta, R. D., & Kundu, D. (1999). Generalized exponential distributions. *Australian and New Zealand Journal of Statistics*, 41(2), 173-188.
- Gupta, R. D., & Kundu, D. (2001a). Generalized exponential distribution: Different methods of estimation. *Journal of Statistical Computation and Simulation*, 69, 315-338.
- Gupta, R. D., & Kundu, D. (2001b). Generalized exponential distribution: An alternative to gamma or Weibull distribution. *Biometrical Journal*, 43, 117-130.
- Gupta, R. D., & Kundu, D. (2002). Generalized exponential distribution: Statistical inferences. *Journal of Statistical Theory and Applications*, 1, 101-118.
- Johnson, N. L., Kotz, S., & Balakrishnan, N. (1994). *Continuous univariate distribution, Vol. 1* (2nd ed.). New York: John Wiley and Sons.
- Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). *Continuous univariate distribution, Vol. 2* (2nd ed.). New York: John Wiley and Sons.
- Mann, N. R., Schafer, R. E., & Singpurwalla, N. D. (1974). *Methods for statistical analysis of reliability and life time data*. New York: John Wiley and Sons.
- Raiffa, H. & Schlaifer, R. (1961). *Applied statistical decision theory*. Cambridge, MA: Harvard University Press.
- Raqab, M. Z. (2002). Inference for generalized exponential distribution based on record statistics. *Journal of Statistical Planning and Inference*, 104, 339-350.
- Raqab, M. Z., & Ahsanullah, M. (2001). Estimation of the location and scale parameters of the generalized exponential distribution based on order statistics. *Journal of Statistical Computation and Simulation*, 69, 109-124.
- Watson, G. N. (1952). *Treatise on the Theory of Bessel Functions* (2nd ed.). Cambridge, UK: Cambridge University Press.
- Zheng, G. (2002). On the Fishers information matrix in type-II censored data from the exponentiated family. *Biometrical Journal*, 44, 353-357.

Appendix A

Lemma 1

Part 1 If X_i are random variables independently and identically generalized exponentially distributed $GED(\alpha, \lambda)$, with λ known, then $T_i = -\ln(1 - e^{-\lambda x_i}) = \ln(1 - e^{-\lambda x_i})^{-1}$ follows $\text{Expo}(\alpha)$.

Part 2 Prove that $\frac{1}{T'}$ follows the inverted gamma distribution, that is $\frac{1}{T'} = y$ follows inverted gamma density (Raiffa & Schlaifer, 1961) as

$$f(y) = \frac{1}{\alpha \Gamma r} e^{\frac{\alpha}{y}} \left(\frac{\alpha}{y} \right)^{r+1} ; y \geq 0, \alpha, r \geq 0 \quad (9)$$

$$E\left(\frac{1}{T'}\right) = \frac{\alpha}{r-1} \text{ and } E\left(\frac{1}{T'}\right)^2 = \frac{\alpha^2}{(r-1)(r-2)} \text{ then } V\left(\frac{1}{T'}\right) = \frac{\alpha^2}{(r-1)^2(r-2)}$$

where $T' = -\sum_{i=1}^n \ln(1 - e^{-\lambda x_i}) = \sum_{i=1}^n \ln(1 - e^{-\lambda x_i})^{-1}$.

LIFE TESTING ANALYSIS OF FAILURE CENSORED DATA

Tables 1 and 2

Table 1. Estimate of parameter α and reliability using the MLE for various known values of λ

λ	$\hat{\alpha}$	$S.E(\hat{\alpha})$	$\hat{R}(t)$	$E(\hat{R}(t))$	$S.E(\hat{R}(t))$
0.01	0.03697	0.01452	0.68107	0.82608	0.08094
0.02	0.04362	0.01714	0.70618	0.82435	0.08484
0.03	0.04872	0.01914	0.72496	0.82303	0.08778
0.04	0.05309	0.02086	0.74074	0.82192	0.09027
0.05	0.05704	0.02241	0.75471	0.82091	0.09248
0.06	0.06071	0.02385	0.76745	0.81999	0.07015
0.07	0.06418	0.02521	0.77928	0.81912	0.07197
0.08	0.06750	0.02652	0.79041	0.81830	0.07369
0.09	0.07070	0.02778	0.80099	0.81751	0.07533
0.10	0.07382	0.02900	0.81111	0.81675	0.07690
0.12	0.07987	0.03137	0.83027	0.81530	0.07989
0.14	0.08573	0.03368	0.84830	0.81391	0.08273
0.15	0.08862	0.03481	0.85697	0.81324	0.08409
0.18	0.09716	0.03817	0.88194	0.81127	0.08805
0.20	0.10279	0.04038	0.89786	0.80999	0.09059
0.30	0.13107	0.05149	0.90123	0.80388	0.06835
0.40	0.16051	0.06305	0.90110	0.79796	0.07581
0.50	0.19190	0.07539	0.85735	0.79209	0.08300
0.60	0.22575	0.08868	0.85331	0.78618	0.09004
0.70	0.26245	0.10310	0.84784	0.78022	0.09695
0.80	0.30232	0.11876	0.84001	0.77419	0.07782
0.90	0.34569	0.13580	0.82882	0.76812	0.08284
1.00	0.39284	0.15432	0.80001	0.76201	0.08778
2.00	1.12735	0.44286	0.71003	0.70934	0.01114
3.00	2.42090	0.95102	0.63922	0.70982	0.00853
4.00	4.15967	1.63407	0.60843	0.59425	0.10358
5.00	6.12946	2.40788	0.45982	0.63671	0.23600
10.00	19.33348	7.59491	0.24577	0.39212	0.20864
20.00	90.37409	35.50230	0.24575	0.35531	0.28917
30.00	338.40470	132.93700	0.24565	0.29058	0.29601

Table 2. Estimate of parameter α and reliability using the UMVUE for various known values of λ

λ	$\tilde{\alpha}$	$S.E(\tilde{\alpha})$	$\tilde{R}(t)$	$E(\tilde{R}(t))$	$S.E(\tilde{R}(t))$
0.01	0.03327	0.01176	0.53961	0.52654	0.08460
0.02	0.03926	0.01388	0.56303	0.54956	0.08372
0.03	0.04385	0.01550	0.58055	0.56679	0.08306
0.04	0.04778	0.01689	0.59529	0.58126	0.08251
0.05	0.05134	0.01815	0.60834	0.59408	0.08202
0.06	0.05464	0.01932	0.62023	0.60577	0.08157
0.07	0.05776	0.02042	0.63129	0.61662	0.08115
0.08	0.06075	0.02148	0.64169	0.62684	0.08076
0.09	0.06363	0.02250	0.65158	0.63655	0.08039
0.10	0.06644	0.02349	0.66104	0.64584	0.08004
0.12	0.07188	0.02541	0.67896	0.66344	0.07936
0.14	0.07716	0.02728	0.69583	0.68000	0.07873
0.15	0.07975	0.02820	0.70396	0.68797	0.07842
0.18	0.08744	0.03091	0.72733	0.71091	0.07755
0.20	0.09251	0.03271	0.74224	0.72553	0.07699
0.30	0.11796	0.04171	0.81143	0.79338	0.07439
0.40	0.14446	0.05107	0.87503	0.85569	0.07200
0.50	0.17271	0.06106	0.93523	0.91465	0.06974
0.60	0.20318	0.07183	0.94094	0.97121	0.06757
0.70	0.23620	0.08351	0.94992	0.90141	0.06548
0.80	0.27209	0.09620	0.95738	0.89999	0.06345
0.90	0.31112	0.11000	0.96834	0.89320	0.02318
1.00	0.35356	0.12500	0.95637	0.81830	0.02318
2.00	1.01461	0.35872	0.82882	0.79209	0.02318
3.00	2.17881	0.77033	0.80001	0.78618	0.02318
4.00	3.74370	1.32360	0.61843	0.78022	0.02318
5.00	5.51652	1.95038	0.49867	0.43011	0.02318
10.00	17.40010	6.15188	0.36759	0.42673	0.02318
20.00	81.33660	28.75680	0.32793	0.33867	0.02318
30.00	304.56400	107.67900	0.27546	0.31526	0.02318

Some Methods of Estimation from Censored Samples in Exponential and Gamma Models

R. R. L. Kantam

Acharya Nagarjuna University
Guntur, India

B. Sriram

ANU College of Engineering & Technology
Guntur, India

Two popular life testing models exponential and one where its generalization is gamma are considered. Estimation of scale parameter from a general Type-II doubly censored sample is attempted by the principle of maximum likelihood method. Resulting equations found to be giving iterative solutions. As an alternative to iterative solution certain admissible modifications to the estimating equations are suggested in special cases. The resulting estimates are compared with the exact maximum likelihood estimates analytically or through simulation. The results are also extended for reliability estimation.

Keywords: Maximum likelihood estimation, Type-II doubly censored sample, order statistics, modified maximum likelihood estimate, asymptotic variance, BLUE, left censored sample, reliability estimation

Introduction

Exponential distribution is the distribution of a continuous life time random variable and is the focal distribution of any reliability study just as normal distribution in classical statistical inference. The gamma distribution is an extension of exponential distribution and its shape parameter θ is more than one. Such an IFR model specifying $\theta = 2$ is considered so that the resulting density is a weighted exponential distribution/length biased version of exponential distribution. The probability density function, the cumulative distribution function, the reliability function, the failure rate of such a distribution are respectively

$$f(z) = ze^{-z} \quad (1)$$

$$F(z) = 1 - e^{-z}(1 + z) \quad (2)$$

R. R. L. Kantam is a professor in the Department of Statistics. Email him at kantam.rrl@gmail.com. B. Sriram is in the Department of Science & Humanities. Email him at sriram_stat@rediffmail.com.

$$R(z) = e^{-z}(1+z) \quad (3)$$

$$h(z) = \frac{z}{1+z} \quad (4)$$

If a scale parameter σ is introduced these functions become respectively

$$f(x) = \frac{x}{\sigma^2} e^{-\frac{x}{\sigma}} \quad (5)$$

$$F(x) = 1 - e^{-\frac{x}{\sigma}} \left(1 + \frac{x}{\sigma}\right) \quad (6)$$

$$R(x) = e^{-\frac{x}{\sigma}} \left(1 + \frac{x}{\sigma}\right) \quad (7)$$

$$h(x) = \frac{x}{\sigma \left(1 + \frac{x}{\sigma}\right)} \quad (8)$$

Generally the life time sample data are generated through a life testing experiment of a sample items giving rise to observations as soon as a sample items fail. The observations flow in a natural order giving rise to ordered sample data (the weaker items fail earlier, stronger items later). In view of the nature of a life testing experiment, it is possible to sometimes have an incomplete data with missing observations on either extremes of the sample – because of very early failures before data collection itself or because some items not failing at all even after a long waiting. Such a sample is called a doubly Type II censored sample. If n items are originally put to test in which it was not possible to have data on lives of the weakest r_1 items, strongest r_2 the sample would be as follows

$$[r_1 \text{ missing}] \ x_{r_1} < x_{r_1+1} < \dots < x_{n-r_2} \quad [r_2 \text{ missing}]$$

$$1, 2, 3, \dots, r_1, r_1 + 1, \dots, n - r_2, n - r_2 + 1, \dots, n$$

This paper deals with the problem of estimation of σ and reliability function $R(x; \sigma)$ of a gamma distribution from a doubly Type II censored samples.

Estimation of σ in a gamma distribution using the well-known ML method is attempted by many researchers from complete as well as censored samples. Whenever a particular method of estimation is resulting in iterative solutions as values of the estimator in any density some admissible adhoc modifications are suggested to that method of estimation by many researchers in order to overcome iterative solution and get a reasonably efficient estimator. In this context Tiku (1967), Mehrotra and Nanda (1974), Cohen and Whitten (1982), Balakrishanan and Cohen (1990), Tiku and Suresh (1992), Rosaiah et al. (1993a), Rosaiah et al. (1993b), Rosaiah et al. (1997), Kantam and Srinivasa Rao (2002), Srinivasa Rao and Kantam (2004), Kantam et al. (2007), Rosaiah et al. (2007) and the references therein are some instances.

Motivated by the modified methods of estimation reported in the literature some modifications are proposed to estimate the scale parameter σ of gamma distribution from a doubly Type II censored sample starting with the well-known maximum likelihood method of estimation. Later modified estimates are used to obtain the reliability estimates also, in addition an attempt is made to establish the asymptotic equivalence of the proposed modifications. Small sample comparisons are also presented. Results of exponential distribution are also given. The findings about estimation of σ are described in the following section and those of reliability estimation are given in the section with that heading.

Parametric Estimation

Let $X_{r_1+1} < X_{r_1+1} < \dots < X_{n-r_2}$ be a doubly Type II censored sample from a scale gamma distribution out of an originally planned random sample of size n in which the least r_1 observations, the largest r_2 observations are missing. The likelihood function of such a sample is given by

$$L \propto [F(x_{r_1+1})]^{r_1} \left[\prod_{i=r_1+1}^{n-r_2} f(x_i) \right] [1-F(x_{n-r_2})]^{r_2} \quad (9)$$

Where $f(\cdot)$, $F(\cdot)$ are given in Equations (5), (6). Substituting the relevant expressions, taking logarithms, differentiating with respect to σ and equating to zero, after some simplifications, results in

$$\frac{-r_1 z_{r_1+1}^2 e^{-z_{r_1+1}}}{1 - (1 + z_{r_1+1}) e^{-z_{r_1+1}}} - 2(n - r_1 - r_2) - \sum_{i=r_1+1}^{n-r_2} z_i - \frac{r_2 z_{n-r_2}^2}{1 + z_{n-r_2}} = 0 \quad (10)$$

Where $z_i = \frac{x_i}{\sigma}$, $i = r_1 + 1, \dots, n - r_2$.

It can be observed that Equation (10) must be solved iteratively to obtain the MLE of σ . However the complete sample situation ($r_1=r_2=0$) gives $\sigma^\wedge = \frac{\bar{x}}{2}$. The failure censored sample situation ($r_1=0, r_2=r$) simplifies Equation (10) to the form

$$A\sigma^2 + B\sigma + C = 0 = \quad (11)$$

$$\left. \begin{aligned} \text{where } A &= 2(n-r) \\ B &= 2(n-r)x_{n-r} - \sum_{i=1}^{n-r} x_i \\ C &= \sum_{i=1}^{n-r} x_i \cdot x_{n-r} + r \cdot x_{n-r}^2 \end{aligned} \right\} \quad (12)$$

It can be proved that the quadratic Equation (11) will have only one positive root and hence is the maximum likelihood estimate of σ , which is

$$\sigma^\wedge = \frac{-B + \sqrt{B^2 - 4AC}}{2A} \quad (13)$$

The exact variance of MLE in Equation (13) cannot be obtained analytically and has to be computed only through simulation and the results were given in Table 4 (all Tables can be found in Appendix A). The asymptotic variance is given by formula

$$\text{asvar}(\sigma^\wedge) = \left[-E \left(\frac{\partial^2 \log L}{\partial \sigma^2} \right) \right]^{-1} \quad (14)$$

The left censored sample situation ($r_1 = r, r_2 = 0$) is of some interest. The estimating equation in this situation becomes

$$\frac{-r z_{r+1}^2 e^{-z_{r+1}}}{1 - [1 + z_{r+1}] e^{-z_{r+1}}} - 2(n-r) - \sum_{i=r+1}^n z_i = 0 \quad (15)$$

This equation cannot be solved for σ analytically because of the structure of the term $\frac{r z_{r+1}^2 e^{-z_{r+1}}}{1 - [1 + z_{r+1}] e^{-z_{r+1}}}$ corresponding to the left censored portion of the sample.

Using the fact that $\frac{x_i}{\sigma} = z_i$ and Equations (5), (6) results in

$$\frac{r z_{r+1}^2 e^{-z_{r+1}}}{1 - [1 + z_{r+1}] e^{-z_{r+1}}} = r z_{r+1} \frac{f(z_{r+1})}{F(z_{r+1})} \quad (16)$$

Unless the portion given by (16) of the Equation (15) is approximated by a reasonable simpler expression, Equation (15) will not give analytical solution for σ to get the MLE from a left censored sample. Similar approximations and modifications are suggested by many authors in many densities earlier. Among such works Tiku (1967) Balakrishnan and Cohen (1990, Ch.6), Tiku and Suresh (1992) are of relevance in the present investigation. Apart from these, specifically for gamma distribution, Mehrotra and Nanda (1974) suggested that the expression in (16) be approximated by its expected value to get simpler estimates of σ . Their approximation resulted in the well-known trimmed mean adjusted for unbiasedness as the approximate maximum likelihood estimator of σ .

The commonness in the modifications of Tiku (1967), Balakrishnan and Cohen (1990) and Tiku and Suresh (1992) is that they have approximated the expression $\frac{f(z_{r+1})}{F(z_{r+1})}$ of Equation (16) by a linear expression of the form $\alpha + \beta z$ (say), where α and β are to be suitably found. This investigation assesses the strength of linearity of the two functions, namely

$$h_1(z) = \frac{f(z_{r+1})}{F(z_{r+1})} = \frac{z e^{-z}}{1 - e^{-z}(1+z)} \quad (17)$$

and

$$h_2(z) = \frac{z_{r+1} f(z_{r+1})}{F(z_{r+1})} = \frac{z^2 \cdot e^{-z}}{1 - e^{-z} (1 + z)} \quad (18)$$

in neighborhoods of $r+1^{\text{th}}$ quantile of the population. Let $p = \frac{r+1}{n+1}$, define

$f(\xi_{r+1}) = \frac{r+1}{n+1}$ so that ξ_r is the r^{th} quantile. Let

$$\left. \begin{aligned} P' &= p - \sqrt{\frac{pq}{n}} \\ P'' &= p + \sqrt{\frac{pq}{n}} \end{aligned} \right\} \quad (19)$$

Where $q=1-p$. Let (z', z'') be the solutions of the equations $F(z') = P'$ and $F(z'') = P''$ where $F(z)$ is given by Equation (6). Thus (z', z'') is an interval in which ξ_{r+1} is expected to fall. For this purpose the values of $F(z)$ for $z = 0$ (0.01) to 11.89 have been tabulated and these are given in Table 1. For each combination of $n=3(1)10$, $r=1,2,\dots,n-2$, identified from Table 1 the values of z close to $F(z')$, $F(z'')$ and supposed to contain the interval (z', z'') . For example, when $n = 5$, $r = 1$:

$$p = \frac{1}{3}, q = \frac{2}{3}$$

$$\begin{aligned} p - \sqrt{\frac{pq}{n}} &= 0.1225, \quad p + \sqrt{\frac{pq}{n}} = 0.54413 \\ F(z') &= 0.1225, \quad F(z'') = 0.54413 \end{aligned}$$

From Table 1, search for the values of $F(z)$ closest to 0.1225 and less than 0.1225, closest to 0.54413 and more than 0.54413. In Table 1 when $z = 0.6$, $F(z) = 0.121901$, when $z = 0.61$, $F(z) = 0.125205$ we take $z^* = 0.6$. Similarly when $z = 1.82$, $F(z) = 0.543087$, when $z = 1.83$ $F(z) = 0.54603$. We take $z^{**} = 1.83$. By the non-decreasing property of distribution function the interval (z', z'') falls in the interval (z^*, z^{**}) .

Table 1 was utilized to obtain the intervals of the type (z^*, z^{**}) for $n = 3(1)10$ $r=1,2,\dots,n-2$. The functions $h_1(z)$ and $h_2(z)$ given by Equations (17), (18) respectively are then evaluated at various values of z over (z^*, z^{**}) . The Karl

Pearson's product moment correlation coefficient is computed between $(z, h_1(z))$, $(z, h_2(z))$ for $z \in (z^*, z^{**})$ with the chosen combinations of n & r . These are given in Table 2.

From Table 2 it may be observed that ρ_2 is larger than ρ_1 uniformly. Therefore it may be concluded that linearization of $h_2(z)$ is preferable to linearization of $h_1(z)$ for $z \in (z^*, z^{**})$. Because (z', z'') is subset of (z^*, z^{**}) we approximate the expression

$$h_2(z_{r+1}) = \frac{e^{-z_{r+1}} z_{r+1}^2}{1 - (1 + z_{r+1}) e^{-z_{r+1}}} \quad (20)$$

of Equation (16) by a linear equation.

i.e.,

$$\frac{e^{-z_{r+1}} z_{r+1}^2}{1 - (1 + z_{r+1}) e^{-z_{r+1}}} \cong \alpha + \beta z_{r+1} \quad (21)$$

where α and β are to be suitably found. Substituting (21) in (15) and solving it for σ results in the modified maximum likelihood estimate of σ from left censored sample as

$$\hat{\sigma} = \frac{\sum_{i=r+2}^n x_i + (1 - r\beta)x_{r+1}}{[r\alpha + 2(n - r)]} \quad (22)$$

This is a linear estimator and its variance can be found using the moments and cross moments of standard ordered gamma statistics given by Prescott (1974), Balasooriya (1992), provided values of α and β are known. Three different methods for calculating α , β are now considered. The basic works relevant to these three methods are – Tiku (1967), Balakrishnan and Cohen (1990, Ch.6), Tiku and Suresh (1992). The methods of getting α and β are described using these works.

Here it should be reemphasized that linearization by earlier authors is for $h_1(z)$, whereas the linearization herein is for $h_2(z)$, because $h_2(z)$ was found to be more linear than $h_1(z)$ as evidenced from sample Table 2.

Method I:

Consider the equations $F(z') = P'$, $F(z'') = P''$ where $p = \frac{r+1}{n+1}$, $q = 1-p$, $P' = p - \sqrt{\frac{pq}{n}}$, $P'' = p + \sqrt{\frac{pq}{n}}$. The values of (z', z'') can be obtained from Table 2.1 for a given, z , n . The values of α and β are given by

$$\beta = \frac{F(z'') - F(z')}{z'' - z'} \quad (23)$$

$$\alpha = h_2(z'') - \beta z'' \text{ or } h_2(z') - \beta z' \quad (24)$$

The values of α , β for $n=3(1)10$ and $r = 1, 2, \dots, n-2$, were calculated and are presented in Table 3. With a specific choice of n and r the modified maximum likelihood estimate by this method is given by Equation (22), Let it be denoted by $\hat{\sigma}_1$. The bias, variance and mean square error of $\hat{\sigma}_1$ are calculated using the moments of standard gamma order statistics, given by Gupta (1960), Prescott (1974). These are given in Table 5.

Method II:

In this method Taylor's series expansion of $h_2(z)$ in the neighborhood of $(r+1)^{\text{th}}$ quantile of the population up to the first derivative is considered, so that $h_2(z_{r+1})$ is approximately equal to $\alpha + \beta z_{r+1}$ where

$$\beta = h'_2(\xi_{r+1}) \quad (25)$$

Here ξ_{r+1} is the solution of $F(\xi_{r+1}) = \frac{r+1}{n+1}$.

$$\alpha = h_2(\xi_{r+1}) - \beta \xi_{r+1} \quad (26)$$

Calculating ξ_{r+1} by interpolation from Table 1 for $n=3(1)10$ and $r = 1, 2, \dots, n-2$, which in turn are used in Equations (25), (26) to obtain α and β . These values of α and β can be used in Equation (22) to get another linear estimator as modified

maximum likelihood estimator say $\hat{\sigma}_2$. For the chosen combinations of n and r the values of α and β by this method are given in [Table 3](#). As described in [Method I](#) the bias, variance and mean square error of $\hat{\sigma}_2$ are given in [Table 5](#).

Method III:

In this method the Taylor's series expansion of $h_2(z)$ in the neighborhood of expected value of $(r+1)^{\text{th}}$ standard order statistic, up to first derivative say e_{r+1} i.e., $E(z_{r+1}) = e_{r+1}$ was considered. If the linear approximation is denoted by the same notation as that of [Methods I](#) and [II](#) without any scope for confusion results in

$$h_2(z_{r+1}) \cong \alpha + \beta z_{r+1}$$

where

$$\beta = h_2'(e_{r+1}) \quad (27)$$

and

$$\alpha = h_2(e_{r+1}) - \beta(e_{r+1}) \quad (28)$$

Using the moments of gamma order statistics given by Gupta (1960) α and β for $n=3(1)10$ and $r = 1, 2, \dots, n-2$ have been evaluated, these are given in [Table 3](#). Substituting the values of α and β in Equation (22) results in another modified maximum likelihood estimator of $\hat{\sigma}_3$ for example,

$$\hat{\sigma}_3 = \sum_{i=r+1}^n l_i x_i \quad (29)$$

Where

$$l_i = \frac{1-r\beta}{r\alpha + 2(n-r)}, i=r+1.$$

$$= \frac{1}{r\alpha + 2(n-r)}, i=r+2, \dots, n$$

α , β in the above l_i 's are given in [Table 3](#) for a specific choice of r and n . Because $\hat{\sigma}_3$ is also a linear estimator its bias, variance and mean square error can be calculated using the moments of gamma order statistics. These three sampling

characteristics of $\hat{\sigma}_3$ are given in Table 5. For the sake of comparison of $\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3$ which are linear estimators the variance of BLUE from a left censored sample of a gamma distribution borrowed from Kantam et al. (1989) is also included in Table 5.

As a matter of academic interest and comparison between CFR and a specific IFR model the calculations for exponential distribution have also been included. Table 6 gives the α, β values for n and r . Table 7 gives the bias, variance, MSE of $\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3$ respectively.

Asymptotica Variance

The log-likelihood equation to estimate $\hat{\sigma}$ from the left censored sample ($r_1 = r, r_2 = 0$) can be written as

$$\frac{\partial \log L}{\partial \sigma} = h_2(z) - 2(n-r) - \sum_{i=r+1}^n z_i$$

$$\text{where } h_2(z) = \frac{z^2 e^{-z}}{1 - e^{-z}(1+z)}.$$

In the suggested three methods of modifications $h_2(z)$ as $\alpha + \beta z$ in the neighborhood of population's $r+1$ quantile or population's mean value of $(r+1)$ sample order statistics were approximated. Because sample quantiles are consistent estimators of population quantiles, sample moments are consistent estimators of population moments, in all the three methods of modification the function $h_2(z)$ is approximated as a linear function of z . For large values of n the neighborhood of population quantile or population mean value becomes narrower thereby giving more linearity of $h_2(z)$ in that neighborhood, that is the closeness of $h_2(z)$ to $\alpha + \beta z$ is stronger, the larger the sample size. Hence for large n , $\frac{\partial \log L}{\partial \sigma}$ is almost equal to $\alpha + \beta z$.

$$\frac{\partial \log L}{\partial \sigma} \cong \frac{\partial \log L'}{\partial \sigma}$$

where

$$\frac{\partial \log L'}{\partial \sigma} = (\alpha + \beta z_{r+1}) - 2(n-r) - \sum_{i=r+1}^n z_i$$

$$\therefore E\left(\frac{\partial \log L}{\partial \sigma}\right) = E\left(\frac{\partial \log L'}{\partial \sigma}\right) \text{ for large } n$$

$$E\left(\frac{\partial^2 \log L}{\partial \sigma^2}\right) = E\left(\frac{\partial^2 \log L'}{\partial \sigma^2}\right) \text{ for large } n$$

MMLEs of [Method I](#), [Method II](#), [Method III](#) have the same asymptotic bias, asymptotic variance as that of MLE. Also BLUE of σ is asymptotically as efficient as MLE and conversely MLEs are asymptotically linear ([David 1981](#), p. 136., p. 273). As such the three MMLEs are asymptotically equivalent to BLUE, exact MLE in the case of left censored samples. For a right censored sample it is the exact MLE and BLUE that stand in competition, which are asymptotically equal.

Small Sample Comparison

The following comparative conclusions from [Table 4](#) are drawn about the performance of exact MLE, BLUE in right censored sample. With respect to variance as well as MSE, it is maximum likelihood estimator that is preferable to BLUE from right censored samples. In case of left censored samples, from [Table 5](#) MMLE under [Method I](#) is the recommended estimator for exponential population also the MMLE under [Method I](#) is most preferable.

Reliability Estimation

The reliability of any item whose life time variate follows gamma distribution is given by

$$R(x) = e^{-\frac{x}{\sigma}} \left(1 + \frac{x}{\sigma}\right) \quad (30)$$

where σ a scale parameter and x is the time instant at which the reliability is desired. If σ is not known reliability is to be estimated through estimators of σ or otherwise Next an attempt is made to estimate reliability using the BLUE, MLE, MMLEs of three methods of [Section 2](#). It is proposed that the corresponding

estimate of reliability with σ in $R(x)$ replaced by $\hat{\sigma}$. This is analogous to the well-known invariance property of ML method of estimation. Though the non ML methods of estimation in this discussion are not established to have invariance property it is exploited here as per remark in Sinha (1986). Accordingly four additional estimators of reliability corresponding to the four methods of estimation of σ are obtained. For one such estimator of reliability the asymptotic variance is given by

$$as\ var\ \hat{R}(x; \sigma) = \left(\frac{\partial R}{\partial \sigma} \right)^2 \cdot asvar(\hat{\sigma}) \quad (31)$$

where $as\ var(\hat{\sigma})$ is the asymptotic variance of $\hat{\sigma}$ obtained by a particular method. For a gamma distribution with shape 2 and scale parameter σ , Equation (31) becomes

$$as\ var\ \hat{R}(z) = \frac{z^4 e^{-2z}}{\sigma^2} \cdot asvar(\hat{\sigma}) \quad (32)$$

where $z = x/\sigma$.

Therefore all the five reliability estimates are asymptotically equally efficient as explained in Section 2. At the same time the small sample variances of the reliability estimates are not mathematically tractable. Monte-Carlo simulation method was used to assess the small sample behavior. 3,000 samples of size $n = 3(1)10$ are generated from a gamma distribution. For each sample after ordering all possible right censored situation with $r = 1(1) n-2$ are extracted. For each right censored sample the exact MLE as the unique positive root of quadratic Equation (11) is computed say $\hat{\sigma}_M$. In order to get the BLUE of σ for the right censored sample the coefficients of uncensored observations for a gamma distribution need to be supplied by a general formula of Lloyd (1952) in scaled densities. As these coefficients are not exhaustively available in published form, they have been evaluated making use of the expected values of Gupta (1960) and variance covariances of Prescott (1974) of gamma order statistics. These coefficients are used to get the BLUE of σ from a right censored sample say $\hat{\sigma}_B$. The gamma reliability function $R(x; \sigma)$ is calculated with σ replaced by $\hat{\sigma}_M, \hat{\sigma}_B$ in succession at values of x corresponding to $R(x) = 0.1(0.1)0.9$. The bias, the variance and MSE of the estimated reliabilities for a given x, n, r a method of estimation across the

3,000 simulated runs are evaluated and these empirical sample characteristic are presented in Table 8.

For left censored samples the exact MLE $\hat{\sigma}_M$ for each simulated sample is obtained as an iterative solution of Equation (15) by Newton-raphson method. The BLUE $\hat{\sigma}_B$ is computed as narrated in the case of right censored samples with the distinction that the weights are calculated for the suffixes of available observations. The three modified maximum likelihood estimators $\hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3$ of the Methods I, II, III respectively are also calculated for each simulated sample using the constants of linearization given in Table 3 in Equation (22). The reliability estimate of left censored sample is obtained in five different ways by successively replacing σ in $R(x; \sigma)$ with $\hat{\sigma}_M, \hat{\sigma}_B, \hat{\sigma}_1, \hat{\sigma}_2, \hat{\sigma}_3$ at values of x corresponding to $R(x) = 0.1(0.1)0.9$ for all left censored samples with $n = 3(1)10, r = 1(1)n-2$. The empirical bias, variance, MSE of all the reliability estimates across the 3,000 simulation runs are presented in Table 9 for $R(x)=0.9$ only a byproduct this entire work is carried out for exponential distribution also in the case of left censored sample only. These results are given in Table 10.

The competition of rating for preferability fell between the use of MLE/BLUE to estimate the gamma reliability function from right censored samples. In the case of left censored samples Method I or Method II of MMLE are therefore suggested. In the case of single parameter exponential distribution the use of BLUE is suggested, although the use of MML Method I or the MML Method II are equally good.

References

- Balakrishnan, N., & Cohen, A. C. (1990). *Order statistics and inference*. London: Academic Press.
- Balasooriya, U., & Hapuaracchihi (1992). Extended tables for the moments of gamma distribution order statistics. *IEEE Transactions on Reliability*, 41(2).
- Cohen, A. C., & Whitten, B. (1982). Modified moment and maximum likelihood estimators for parameters of the three-parameter gamma distribution. *Commun. Statist.-Simu. Comp.*, 11, 197-216.
- David, H. A. (1981). *Order statistics*. New York: John Wiley and Sons.
- Gupta, S. S. (1960). Order statistics from the gamma distribution. *Technometrics*, 2, 243-262.

- Kantam, R. R. L., Anjaneyulu, G. V. S. R., & Narasimham, V.L. (1989). On linear estimation of gamma scale parameter from censored samples. *Assam Statist.Rev.*, 3(1), 1-21.
- Kantam, R. R. L., & Srinivasa, R. G. (2002). Log-Logistic distribution: Modified Maximum Likelihood estimation. *Gujarat Statistical Review*, 29(1-2), 25-36.
- Kantam, R. R. L., Srinivasa, R. G., & Sriram,B. (2007). On stress-strength reliability model in log-logistic distribution. *International Journal of Agricultural Statistical Science*, 3(1), 1-6.
- Lloyd, E. H. (1952). Least-squares estimation of location and scale parameters using order statistics. *Biometrika*, 39, 88-95.
- Mehrotra, K. G., & Nanda, P. (1974). Unbiased estimation of parameters by order statistics in the case of censored samples. *Biometrika*, 61, 601-606.
- Prescott, P. (1974). Variances and covariances of order statistics from the gamma distribution. *Biometrika*, 61, 607-613.
- Rosaiah, K., Kantam, R. R. L. & Narasimham, V. L. (1993a). ML and modified ML estimation in gamma distribution with a known prior relation among the parameters. *Pakistan Journal of Statistics*, 9(3), B, 37-48.
- Rosaiah, K. Kantam, R. R. L., & Narasimham, V. L. (1993b). On modified maximum likelihood estimation of gamma parameters. *Journal of Statistical Research, Bangladesh*, 27(1-2), 15-28.
- Rosaiah, K., Kantam, R. R. L., & Jayaprakash, B. (1997). Reliability estimation in half logistic model. *Gujarat Statistical Review*, 24(2), 11-18.
- Rosaiah, K., Kantam, R. R. L., & Pratapa Reddy, J. (2007). Estimation in inverse Rayleigh distribution. *International Journal of Agricultural Statistical Science*, 3(2), 453-470.
- Sinha, S. K. (1986). Reliability and life testing. *Wiley Eastern Limited*.
- Srinivasa, R. G. & Kantam, R. R. L. (2004). A note on point estimation of system reliability exemplified for the log-logistic distribution. *Economic Quality Control*, 19(2), 197-204.
- Tiku, M. L. (1967). Estimating the mean and standard deviation from a censored normal sample. *Biometrika*, 54, 155-165.
- Tiku, M. L. & Suresh, R. P. (1992). A new method of estimation for location and scale parameters. *Journal of Statistical Planning and Inference*, 30, 281-292.

Appendix A

Table 1. Values of cumulative distribution function of Gamma distribution with shape parameter 2

Z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.00	0.00000	0.00005	0.00019	0.00044	0.00077	0.00120	0.00173	0.00233	0.00303	0.00381
0.10	0.00467	0.00562	0.00664	0.00775	0.00893	0.01018	0.01151	0.01291	0.01438	0.01591
0.20	0.01752	0.01919	0.02092	0.02272	0.02458	0.02649	0.02847	0.03050	0.03259	0.03474
0.30	0.03693	0.03918	0.04148	0.04383	0.04622	0.04867	0.05116	0.05369	0.05627	0.05889
0.40	0.06155	0.06425	0.06699	0.06977	0.07258	0.07543	0.07832	0.08124	0.08420	0.08918
0.50	0.09020	0.09325	0.09632	0.09943	0.10256	0.10572	0.10891	0.11212	0.11536	0.11862
0.60	0.12190	0.12520	0.12853	0.13187	0.13524	0.13862	0.14202	0.14544	0.14888	0.15233
0.70	0.15580	0.15928	0.16278	0.16629	0.16982	0.17335	0.17690	0.18046	0.18403	0.18761
0.80	0.19120	0.19480	0.19841	0.20203	0.20565	0.20928	0.21291	0.21656	0.22020	0.22386
0.90	0.22751	0.23117	0.23484	0.23851	0.24218	0.24585	0.24953	0.25320	0.25688	0.26056
1.00	0.26424	0.26792	0.27159	0.27527	0.27895	0.28262	0.28630	0.28997	0.29364	0.29730
1.10	0.30097	0.30463	0.30828	0.31193	0.31558	0.31923	0.32287	0.32650	0.33013	0.33375
1.20	0.33737	0.34098	0.34458	0.34818	0.35177	0.35536	0.35894	0.36251	0.36607	0.36963
1.30	0.37317	0.37671	0.38024	0.38376	0.38728	0.39078	0.39428	0.39776	0.40124	0.40471
1.40	0.40816	0.41161	0.41505	0.41847	0.42189	0.42530	0.42869	0.43208	0.43545	0.43882
1.50	0.44217	0.44551	0.44884	0.45216	0.45547	0.45876	0.46205	0.46532	0.46858	0.47183
1.60	0.47506	0.47829	0.48150	0.48470	0.48789	0.49106	0.49423	0.49738	0.50051	0.50364
1.70	0.50675	0.50985	0.51294	0.51601	0.51907	0.52212	0.52515	0.52817	0.53118	0.53418
1.80	0.53716	0.54013	0.54308	0.54603	0.54895	0.55187	0.55477	0.55766	0.56054	0.56340
1.90	0.56625	0.56908	0.57190	0.57471	0.57751	0.58029	0.83059	0.58581	0.58855	0.59128
2.00	0.59399	0.59669	0.59938	0.60205	0.60471	0.60735	0.60999	0.61261	0.61521	0.61780
2.10	0.62038	0.62295	0.62550	0.62803	0.63056	0.63307	0.63557	0.63805	0.64052	0.64298
2.20	0.64543	0.64786	0.65027	0.65268	0.65507	0.65745	0.65981	0.66216	0.66450	0.66683
2.30	0.66914	0.67144	0.67373	0.67600	0.67826	0.68051	0.68274	0.68497	0.68717	0.68937
2.40	0.69155	0.69373	0.69588	0.69803	0.70016	0.70228	0.70439	0.70649	0.70857	0.71064
2.50	0.71270	0.71474	0.71675	0.71880	0.72081	0.72281	0.72479	0.72676	0.72872	0.73067
2.60	0.73261	0.73454	0.73645	0.73835	0.74024	0.74212	0.74399	0.74584	0.74768	0.74951
2.70	0.75134	0.75314	0.75494	0.75673	0.75850	0.76027	0.76202	0.76376	0.76549	0.76721
2.80	0.76892	0.77061	0.77230	0.77398	0.77564	0.77729	0.77894	0.78057	0.78219	0.78380
2.90	0.78540	0.78700	0.78858	0.79015	0.79170	0.79325	0.79479	0.79632	0.79784	0.79935

Table 1 (contd.). Values of cumulative distribution function of Gamma distribution with shape parameter 2

Z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
3.00	0.80085	0.80234	0.80381	0.80528	0.80674	0.80819	0.80963	0.81106	0.81248	0.81389
3.10	0.81529	0.81669	0.81807	0.81944	0.82080	0.82216	0.82350	0.82484	0.82617	0.82749
3.20	0.82879	0.83009	0.83139	0.83267	0.83394	0.83521	0.83646	0.83771	0.83895	0.84018
3.30	0.84140	0.84261	0.84382	0.84501	0.84620	0.84738	0.84855	0.84971	0.85087	0.85201
3.40	0.85315	0.85418	0.85541	0.85652	0.85763	0.85873	0.85982	0.86090	0.86198	0.86305
3.50	0.86411	0.86516	0.86621	0.86724	0.86827	0.86930	0.87031	0.87132	0.87232	0.87332
3.60	0.87431	0.87529	0.87626	0.87723	0.87818	0.87914	0.88008	0.88102	0.88195	0.88288
3.70	0.88379	0.88471	0.88561	0.88651	0.88740	0.88829	0.88916	0.89004	0.89090	0.89176
3.80	0.89262	0.89346	0.89430	0.89514	0.89597	0.89679	0.89761	0.89842	0.89922	0.90002
3.90	0.90082	0.90160	0.90238	0.90316	0.90393	0.90469	0.90545	0.90620	0.90695	0.90769
4.00	0.90842	0.90915	0.90988	0.91060	0.91131	0.91202	0.91272	0.91342	0.91411	0.91480
4.10	0.91548	0.91616	0.91683	0.91750	0.91816	0.91881	0.91947	0.92011	0.92075	0.92139
4.20	0.92202	0.92265	0.92327	0.92389	0.92451	0.92511	0.92572	0.92632	0.92691	0.92750
4.30	0.92809	0.92867	0.92925	0.92982	0.93039	0.93095	0.93151	0.93206	0.93261	0.93316
4.40	0.93370	0.93424	0.93478	0.93531	0.93583	0.93635	0.93687	0.93738	0.93789	0.93840
4.50	0.93890	0.93940	0.93989	0.94038	0.94087	0.94135	0.94183	0.94231	0.94278	0.94325
4.60	0.94371	0.94417	0.94463	0.94508	0.94553	0.94598	0.94642	0.94686	0.94730	0.94773
4.70	0.94816	0.94858	0.94901	0.94943	0.94984	0.95025	0.95066	0.95107	0.95147	0.95187
4.80	0.95227	0.95266	0.95305	0.95344	0.95382	0.95420	0.95458	0.95496	0.95533	0.95570
4.90	0.95607	0.95643	0.95679	0.95715	0.95750	0.95785	0.95820	0.95855	0.95889	0.95923
5.00	0.95957	0.95991	0.96024	0.96057	0.96090	0.96122	0.96155	0.96187	0.96218	0.96250
5.10	0.96281	0.96312	0.96343	0.96373	0.96403	0.96433	0.96463	0.96493	0.96522	0.96551
5.20	0.96580	0.96608	0.96637	0.96665	0.96693	0.96720	0.96748	0.96775	0.96802	0.96829
5.30	0.96855	0.96882	0.96908	0.96934	0.96959	0.96985	0.97010	0.97035	0.97060	0.97085
5.40	0.97109	0.97134	0.97158	0.97182	0.97205	0.97229	0.97252	0.97275	0.97298	0.97321
5.50	0.97344	0.97366	0.97388	0.97410	0.97432	0.97454	0.97475	0.97497	0.97518	0.97539
5.60	0.97559	0.97580	0.97601	0.97621	0.97641	0.97661	0.97681	0.97700	0.97720	0.97739
5.70	0.97758	0.97777	0.97796	0.97815	0.97833	0.97852	0.97870	0.97888	0.97906	0.97924
5.80	0.97941	0.97959	0.97976	0.97993	0.98010	0.98027	0.98044	0.98061	0.98077	0.98094
5.90	0.98110	0.98126	0.98142	0.98158	0.98173	0.98189	0.98204	0.98220	0.98235	0.98250

ESTIMATION FROM SAMPLES IN EXPONENTIAL & GAMMA MODELS

Table 1 (contd.). Values of cumulative distribution function of Gamma distribution with shape parameter 2

Z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
6.00	0.98265	0.98280	0.98294	0.98309	0.98323	0.98338	0.98352	0.98366	0.98380	0.98394
6.10	0.98408	0.98421	0.98435	0.98448	0.98461	0.98475	0.98488	0.98501	0.98513	0.98526
6.20	0.98539	0.98551	0.98564	0.98576	0.98588	0.98600	0.98612	0.98624	0.98636	0.98648
6.30	0.98660	0.98671	0.98682	0.98694	0.98705	0.98716	0.98727	0.98738	0.98749	0.98760
6.40	0.98771	0.98781	0.98792	0.98802	0.98812	0.98823	0.98833	0.98843	0.98852	0.98862
6.50	0.98872	0.98882	0.98892	0.98901	0.98911	0.98920	0.98930	0.98939	0.98948	0.98957
6.60	0.98966	0.98975	0.98984	0.98993	0.99001	0.99010	0.99019	0.99027	0.99036	0.99044
6.70	0.99052	0.99060	0.99069	0.99077	0.99085	0.99093	0.99100	0.99108	0.99116	0.99124
6.80	0.99131	0.99139	0.99146	0.99154	0.99161	0.99168	0.99176	0.99183	0.99190	0.99197
6.90	0.99204	0.99211	0.99218	0.99224	0.99231	0.99238	0.99245	0.99251	0.99258	0.99264
7.00	0.99271	0.99277	0.99283	0.99289	0.99296	0.99302	0.99308	0.99314	0.99320	0.99326
7.10	0.99332	0.99338	0.99343	0.99349	0.99355	0.99360	0.99366	0.99372	0.99377	0.99382
7.20	0.99388	0.99393	0.99399	0.99404	0.99409	0.99414	0.99419	0.99424	0.99429	0.99434
7.30	0.99439	0.99444	0.99449	0.99454	0.99459	0.99463	0.99468	0.99473	0.99477	0.99482
7.40	0.99487	0.99491	0.99496	0.99500	0.99504	0.99509	0.99513	0.99517	0.99522	0.99526
7.50	0.99530	0.99534	0.99538	0.99542	0.99546	0.99550	0.99554	0.99558	0.99562	0.99566
7.60	0.99570	0.99573	0.99577	0.99581	0.99585	0.99588	0.99592	0.99595	0.99599	0.99603
7.70	0.99606	0.99610	0.99613	0.99616	0.99620	0.99623	0.99626	0.99630	0.99633	0.99636
7.80	0.99639	0.99643	0.99646	0.99649	0.99652	0.99655	0.99658	0.99661	0.99664	0.99667
7.90	0.99670	0.99673	0.99676	0.99679	0.99682	0.99684	0.99687	0.99690	0.99693	0.99695
8.00	0.99698	0.99701	0.99703	0.99706	0.99709	0.99711	0.99714	0.99716	0.99719	0.99721
8.10	0.99724	0.99726	0.99729	0.99731	0.99733	0.99736	0.99738	0.99741	0.99743	0.99745
8.20	0.99747	0.99750	0.99752	0.99754	0.99756	0.99758	0.99761	0.99763	0.99765	0.99767
8.30	0.99769	0.99771	0.99773	0.99775	0.99777	0.99779	0.99781	0.99783	0.99779	0.99787
8.40	0.99789	0.99791	0.99792	0.99794	0.99796	0.99798	0.99800	0.99801	0.99803	0.99805
8.50	0.99807	0.99808	0.99810	0.99812	0.99814	0.99815	0.99817	0.99818	0.99820	0.99822
8.60	0.99823	0.99825	0.99826	0.99828	0.99830	0.99831	0.99833	0.99834	0.99836	0.99837
8.70	0.99838	0.99840	0.99841	0.99843	0.99844	0.99846	0.99847	0.99848	0.99850	0.99851
8.80	0.99852	0.99854	0.99855	0.99856	0.99858	0.99859	0.99860	0.99861	0.99863	0.99864
8.90	0.99865	0.99866	0.99867	0.99869	0.99870	0.99871	0.99872	0.99873	0.99874	0.99876

Table 1 (contd.). Values of cumulative distribution function of Gamma distribution with shape parameter 2

Z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
9.00	0.99877	0.99878	0.99879	0.99880	0.99881	0.99882	0.99883	0.99884	0.99885	0.99886
9.10	0.99887	0.99888	0.99889	0.99890	0.99891	0.99892	0.99893	0.99894	0.99895	0.99896
9.20	0.99897	0.99898	0.99899	0.99900	0.99901	0.99902	0.99902	0.99903	0.99904	0.99905
9.30	0.99906	0.99907	0.99908	0.99908	0.99909	0.99910	0.99911	0.99912	0.99912	0.99913
9.40	0.99914	0.99915	0.99916	0.99916	0.99917	0.99918	0.99919	0.99919	0.99920	0.99921
9.50	0.99924	0.99922	0.99923	0.99924	0.99924	0.99925	0.99926	0.99926	0.99927	0.99928
9.60	0.99928	0.99929	0.99930	0.99930	0.99931	0.99931	0.99932	0.99933	0.99933	0.99934
9.70	0.99934	0.99935	0.99936	0.99936	0.99937	0.99937	0.99938	0.99939	0.99939	0.99940
9.80	0.99940	0.99941	0.99941	0.99942	0.99942	0.99943	0.99943	0.99944	0.99944	0.99945
9.90	0.99945	0.99946	0.99946	0.99947	0.99948	0.99948	0.99948	0.99949	0.99949	0.99950
10.00	0.99950	0.99951	0.99951	0.99951	0.99952	0.99952	0.99953	0.99953	0.99954	0.99954
10.10	0.99954	0.99955	0.99955	0.99956	0.99956	0.99956	0.99957	0.99957	0.99958	0.99958
10.20	0.99958	0.99959	0.99959	0.99960	0.99960	0.99960	0.99961	0.99961	0.99961	0.99962
10.30	0.99962	0.99962	0.99963	0.99963	0.99963	0.99964	0.99964	0.99964	0.99965	0.99965
10.40	0.99965	0.99966	0.99966	0.99966	0.99967	0.99967	0.99967	0.99968	0.99968	0.99968
10.50	0.99968	0.99969	0.99969	0.99969	0.99970	0.99972	0.99970	0.99970	0.99971	0.99971
10.60	0.99971	0.99971	0.99972	0.99972	0.99972	0.99975	0.99973	0.99973	0.99973	0.99973
10.70	0.99974	0.99974	0.99974	0.99974	0.99975	0.99977	0.99975	0.99975	0.99976	0.99976
10.80	0.99976	0.99976	0.99976	0.99977	0.99977	0.99979	0.99977	0.99977	0.99978	0.99978
10.90	0.99978	0.99978	0.99978	0.99979	0.99979	0.99981	0.99979	0.99979	0.99980	0.99980
11.00	0.99980	0.99980	0.99980	0.99981	0.99981	0.99983	0.99981	0.99981	0.99981	0.99982
11.10	0.99982	0.99982	0.99982	0.99982	0.99982	0.99984	0.99983	0.99983	0.99983	0.99983
11.20	0.99983	0.99984	0.99984	0.99984	0.99984	0.99986	0.99984	0.99984	0.99985	0.99985
11.30	0.99985	0.99985	0.99985	0.99985	0.99985	0.99987	0.99986	0.99986	0.99986	0.99986
11.40	0.99986	0.99986	0.99986	0.99987	0.99987	0.99987	0.99987	0.99987	0.99987	0.99987
11.50	0.99987	0.99988	0.99988	0.99988	0.99988	0.99988	0.99988	0.99988	0.99988	0.99988
11.60	0.99988	0.99988	0.99989	0.99989	0.99989	0.99989	0.99989	0.99989	0.99989	0.99989
11.70	0.99990	0.99989	0.99990	0.99990	0.99990	0.99989	0.99990	0.99990	0.99990	0.99990
11.80	0.99990	0.99991	0.99991	0.99991	0.99991	0.99991	0.99991	0.99991	0.99991	0.99991

ESTIMATION FROM SAMPLES IN EXPONENTIAL & GAMMA MODELS

Table 2. Gamma (2) distribution correlation coefficients of $(z, h_1(z)), (z, h_2(z)), z \in (z^*, z^{**})$

n	r	z^*	z^{**}	r_1	r_2
5	1	0.6	1.83	0.9528	0.999
	2	1.03	2.56	0.9633	0.99838
	3	1.54	3.64	0.9626	0.99548
10	1	0.39	1.11	0.94753	0.99177
	2	0.63	1.42	0.97025	0.9992
	3	0.86	1.73	0.97219	0.99897
	4	1.09	2.07	0.97798	0.99922
	5	1.34	2.45	0.979	0.9992
	6	1.63	2.92	0.97784	0.99693
	7	1.97	3.54	0.9769	0.99597
	8	2.42	4.53	0.9555	0.98629

Table 3. Gamma (2) distribution constants of linear approximation for MMLEs from left censored sample

<i>n</i>	<i>r</i>	Method I		Method II		Method III	
		<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>
3	1	1.84091	-0.43072	1.81952	-0.45748	1.78597	-0.43832
4	1	1.90320	-0.48750	1.88015	-0.49717	1.86883	-0.48912
	2	1.73474	-0.37991	1.73608	-0.41238	1.68453	-0.38802
5	1	1.93150	-0.51815	1.91158	-0.52169	1.90748	-0.51826
	2	1.83181	-0.44270	1.81953	-0.45749	1.79827	-0.44516
	3	1.64480	-0.34330	1.66186	-0.37794	1.59699	-0.35086
6	1	1.94746	-0.53785	1.93053	-0.53854	1.92936	-0.56374
	2	1.87819	-0.47836	1.86467	-0.48625	1.85453	-0.47943
	3	1.83834	-0.43395	1.76285	-0.42596	1.73350	-0.41111
	4	1.56715	-0.31499	1.59551	-0.35027	1.52114	-0.32191
7	1	1.95763	-0.55182	1.94311	-0.55099	1.94331	-0.55119
	2	1.90588	-0.50233	1.89266	-0.50650	1.88752	-0.50261
	3	1.82782	-0.44717	1.81952	-0.45748	1.80406	-0.44845
	4	1.70687	-0.38095	1.71033	-0.39994	1.67437	-0.38346
	5	1.50449	-0.29485	1.53614	-0.32743	1.45456	-0.29843
8	1	1.96537	-0.56323	1.95270	-0.56142	1.95288	-0.56163
	2	1.92479	-0.52065	1.91298	-0.52285	1.90898	-0.51950
	3	1.86766	-0.47549	1.83779	-0.46865	1.84663	-0.47270
	4	1.77636	-0.41916	1.77284	-0.43123	1.75654	-0.42269
	5	1.64785	-0.35591	1.65752	-0.37605	1.62040	-0.36036
	6	1.43326	-0.27115	1.47800	-0.30651	1.39558	-0.27889
9	1	1.96979	-0.57064	1.95863	-0.56838	1.95988	-0.56989
	2	1.93576	-0.53244	1.92511	-0.53351	1.92395	-0.53246
	3	1.88952	-0.49297	1.88015	-0.49717	1.87488	-0.49338
	4	1.82588	-0.44958	1.81952	-0.45748	1.80735	-0.45034
	5	1.73608	-0.39956	1.73608	-0.41237	1.71209	-0.40077
	6	1.60303	-0.33858	1.61677	-0.35886	1.57737	-0.34312
	7	1.38549	-0.25722	1.32181	-0.29151	1.34407	-0.26268
10	1	1.97368	-0.57735	1.96373	-0.57476	1.96506	-0.57649
	2	1.94494	-0.54287	1.93525	-0.54308	1.93487	-0.5427
	3	1.90697	-0.50797	1.89806	-0.51069	1.89489	-0.50822
	4	1.85793	-0.47124	1.84941	-0.47607	1.84165	-0.47108
	5	1.78869	-0.42895	1.78506	-0.43782	1.76991	-0.42967
	6	1.69455	-0.38063	1.69791	-0.39414	1.67028	-0.38164
	7	1.55743	-0.32171	1.57513	-0.34224	1.52547	-0.32349
	8	1.3371	-0.24336	1.38993	-0.27708	1.29554	-0.24805

ESTIMATION FROM SAMPLES IN EXPONENTIAL & GAMMA MODELS

Table 4. Gamma (2) distribution empirical sample characteristics of MMLEs of σ from right censored sample

n	r	BIAS	VAR		MSE
		EXACT MLE	EXT MLE	BLUE	EXACT MLE
3	1	0.00975	0.22591	0.22803	0.22600
4	1	0.00277	0.15943	0.15510	0.15944
	2	0.01334	0.21390	0.21523	0.21408
5	1	0.00563	0.12188	0.11782	0.12191
	2	0.00281	0.15087	0.14766	0.15088
	3	0.02190	0.19516	0.20653	0.19564
6	1	0.00870	0.09425	0.09509	0.09433
	2	0.00771	0.11231	0.11285	0.11237
	3	0.01401	0.14175	0.14221	0.14194
	4	0.02220	0.18893	0.20008	0.18942
7	1	0.00963	0.07646	0.07975	0.07655
	2	0.00852	0.08823	0.09154	0.08830
	3	0.01057	0.10779	0.10913	0.10790
	4	0.01027	0.13416	0.13807	0.13426
	5	0.02869	0.17353	0.19505	0.17435
8	1	0.00434	0.06593	0.06870	0.06594
	2	0.01160	0.07706	0.07708	0.07720
	3	0.00288	0.08631	0.08878	0.08632
	4	0.00109	0.10277	0.10614	0.10277
	5	0.01748	0.12868	0.13469	0.12899
	6	0.01791	0.19094	0.19103	0.19126
9	1	0.00464	0.06160	0.06035	0.06163
	2	0.00044	0.06707	0.06661	0.06707
	3	0.00442	0.07414	0.07495	0.07416
	4	0.00048	0.08776	0.08654	0.08776
	5	0.00136	0.10196	0.10372	0.10196
	6	0.01426	0.12706	0.13196	0.12726
	7	0.02398	0.17928	0.18767	0.17986
10	1	0.00018	0.05285	0.05382	0.05285
	2	0.00243	0.05726	0.05867	0.05726
	3	0.00086	0.06598	0.06492	0.06598
	4	0.00204	0.07066	0.07319	0.07067
	5	0.00035	0.08212	0.08465	0.08212
	6	0.01394	0.09408	0.10168	0.09428
	7	0.00585	0.12892	0.12963	0.12895
	8	0.03092	0.17385	0.18487	0.17481

Table 5. Gamma (2) distribution empirical sample characteristics of MMLEs of σ from left censored sample

BIAS						VAR					MSE			
n	r	σ_M	σ_1	σ_2	σ_3	σ_M	σ_B	σ_1	σ_2	σ_3	σ_M	σ_1	σ_2	σ_3
3	1	0.00029	0.00312	0.00893	0.00874	0.17905	0.17425	0.17425	0.17427	0.17426	0.17905	0.17629	0.17748	0.17739
4	1	0.01058	0.00088	0.00381	0.00378	0.13410	0.12758	0.12758	0.12757	0.12758	0.13422	0.12735	0.12856	0.12855
	2	0.00624	0.00435	0.01450	0.01408	0.13459	0.13744	0.13745	0.13749	0.13746	0.13463	0.13628	0.14172	0.14156
5	1	0.00050	0.00040	0.00204	0.00203	0.10704	0.10115	0.10118	0.10323	0.12016	0.10704	0.10110	0.10366	0.12065
	2	0.00016	0.00120	0.00670	0.00668	0.10403	0.10489	0.10491	0.11263	0.11227	0.10403	0.10466	0.11420	0.11382
	3	0.02087	0.00510	0.01820	0.01759	0.12203	0.11552	0.11547	0.12794	0.12668	0.12247	0.11433	0.13296	0.13148
6	1	0.01104	0.00023	0.00125	0.00125	0.08554	0.08394	0.08394	0.08394	0.08394	0.08566	0.08390	0.08415	0.08414
	2	0.00055	0.00053	0.00381	0.00378	0.08389	0.08575	0.08575	0.08575	0.08575	0.08389	0.08566	0.08643	0.08641
	3	0.00691	0.00409	0.00891	0.00876	0.09437	0.09008	0.08963	0.09009	0.09008	0.09442	0.08891	0.09178	0.09175
	4	0.01417	0.00568	0.02067	0.01998	0.10788	0.10087	0.10088	0.10093	0.10089	0.10808	0.09976	0.10557	0.10536
7	1	0.00943	0.00015	0.00083	0.00083	0.07002	0.07178	0.07179	0.07178	0.07178	0.07011	0.07178	0.07190	0.07190
	2	0.00152	0.00031	0.00241	0.00241	0.07475	0.07279	0.07278	0.07278	0.07278	0.07475	0.07274	0.07314	0.07314
	3	0.00479	0.00063	0.00523	0.00518	0.07750	0.07497	0.07497	0.07498	0.07497	0.07753	0.07488	0.07579	0.07578
	4	0.00659	0.00163	0.01057	0.01036	0.08170	0.07958	0.07958	0.07959	0.07958	0.08174	0.07932	0.08140	0.08134
	5	0.01722	0.00495	0.02249	0.02174	0.09591	0.09093	0.09034	0.09039	0.09052	0.09620	0.08947	0.09501	0.09479

ESTIMATION FROM SAMPLES IN EXPONENTIAL & GAMMA MODELS

Table 5 (contd.). Gamma (2) distribution empirical sample characteristics of MMLEs of σ from left censored sample

		BIAS				VAR					MSE			
<i>n</i>	<i>r</i>	σ_M	σ_1	σ_2	σ_3	σ_M	σ_B	σ_1	σ_2	σ_3	σ_M	σ_1	σ_2	σ_3
8	1	0.00487	0.00010	0.00059	0.00059	0.06025	0.06273	0.06127	0.06273	0.06273	0.06027	0.06126	0.06280	0.06280
	2	0.00484	0.00017	0.00165	0.00165	0.06430	0.06393	0.06333	0.06334	0.06333	0.06432	0.06331	0.06355	0.06355
	3	0.00107	0.00030	0.00342	0.00339	0.06424	0.06459	0.06459	0.06461	0.06459	0.06424	0.06455	0.06504	0.06504
	4	0.00773	0.00079	0.00636	0.00196	0.06513	0.06699	0.06699	0.06700	0.06758	0.06519	0.06683	0.06789	0.06785
	5	0.00540	0.00193	0.01179	0.01161	0.07190	0.07173	0.07173	0.07175	0.07173	0.07193	0.07145	0.07359	0.07355
	6	0.02547	0.00672	0.02374	0.02308	0.08639	0.08295	0.08235	0.08241	0.08236	0.08704	0.08129	0.08693	0.08674
9	1	0.00389	0.00008	0.00044	0.00044	0.05602	0.05571	0.05571	0.05571	0.05571	0.05604	0.05569	0.05576	0.05575
	2	0.00179	0.00013	0.00119	0.00119	0.05655	0.05610	0.05610	0.05610	0.05610	0.05655	0.05608	0.05624	0.05624
	3	0.00049	0.00021	0.00238	0.00238	0.05533	0.05689	0.05689	0.05689	0.05689	0.05533	0.05686	0.05716	0.05716
	4	0.00800	0.00039	0.00425	0.00422	0.05885	0.05891	0.05831	0.05831	0.05831	0.05891	0.05826	0.05883	0.05882
10	5	0.00482	0.00077	0.00733	0.00723	0.06151	0.06084	0.06085	0.06085	0.06085	0.06154	0.06075	0.06180	0.06178
	6	0.00731	0.00191	0.01288	0.01258	0.06874	0.06563	0.06563	0.06565	0.06563	0.06880	0.06538	0.06752	0.06746
	7	0.02974	0.00662	0.08570	0.02413	0.08059	0.07607	0.07607	0.07613	0.07608	0.08147	0.07512	0.09708	0.08038
	1	0.00178	0.00003	0.00036	0.00036	0.04805	0.05011	0.05010	0.05059	0.05059	0.04805	0.05010	0.05063	0.05063
	2	0.00181	0.00005	0.00095	0.00095	0.04765	0.05038	0.05036	0.05258	0.05258	0.04766	0.05036	0.05268	0.05268

Table 6. Exponential distribution constants of linear approximation for MMLEs from left censored sample

n	r	Method I		Method II		Method III	
		α	β	α	β	α	β
3	1	0.97063	-0.35651	0.96090	-0.38629	0.94408	-0.36425
4	1	0.98576	-0.40146	0.97853	-0.41559	0.97211	-0.40386
	2	0.93657	-0.31895	0.93287	-0.35143	0.90767	-0.32622
5	1	0.99154	-0.42443	0.98640	-0.43278	0.98329	-0.42550
	2	0.96653	-0.36966	0.96091	-0.38629	0.95039	-0.37206
	3	0.90325	-0.29059	0.90521	-0.32396	0.87336	-0.29720
6	1	0.99438	-0.43859	0.99062	-0.44413	0.98887	-0.43916
	2	0.97914	-0.39740	0.97431	-0.40769	0.96890	-0.39851
	3	0.94575	-0.34405	0.94226	-0.36208	0.92807	-0.34628
	4	0.87203	-0.26813	0.87888	-0.30155	0.84178	-0.27408
7	1	0.99600	-0.44824	0.99313	-0.45218	0.99166	-0.44857
	2	0.98570	-0.41520	0.98179	-0.42224	0.97864	-0.41581
	3	0.96487	-0.37474	0.96090	-0.38629	0.95328	-0.37579
	4	0.92481	-0.32278	0.92354	-0.34159	0.90617	-0.32484
	5	0.85945	-0.26815	0.85414	-0.28279	0.81291	-0.25510
8	1	0.99711	-0.45577	0.99487	-0.45867	0.99404	-0.45546
	2	0.98988	-0.42856	0.98674	-0.43361	0.98441	-0.42803
	3	0.97598	-0.39663	0.97245	-0.40443	0.96711	-0.39564
	4	0.95142	-0.35785	0.94852	-0.36968	0.93727	-0.35632
	5	0.90254	-0.30299	0.90357	-0.32247	0.88505	-0.30655
	6	0.81412	-0.23256	0.82901	-0.26546	0.78652	-0.23915
9	1	0.99767	-0.46057	0.99586	-0.46287	0.99537	-0.46072
	2	0.99206	-0.43689	0.98947	-0.44081	0.98812	-0.43714
	3	0.98165	-0.40959	0.97853	-0.41559	0.97555	-0.40995
	4	0.96397	-0.37744	0.96091	-0.38629	0.95493	-0.37798
	5	0.93461	-0.33850	0.93287	-0.35143	0.92133	-0.33935
	6	0.88472	-0.28916	0.88748	-0.30855	0.86495	-0.29076
10	7	0.79223	-0.22112	0.80946	-0.25295	0.76233	-0.22551
	1	0.99815	-0.46478	0.99666	-0.46662	0.99630	-0.46488
	2	0.99374	-0.44402	0.99158	-0.44710	0.99063	-0.44420
	3	0.98582	-0.42056	0.98315	-0.42518	0.98111	-0.42079
	4	0.97280	-0.39354	0.96993	-0.40019	0.96597	-0.39387
	5	0.95220	-0.36188	0.94976	-0.37125	0.94247	-0.36234
	6	0.91952	-0.32366	0.91892	-0.33695	0.90570	-0.32440
	7	0.86594	-0.27553	0.87044	-0.29494	0.84590	-0.27695
	8	0.76974	-0.20972	0.78932	-0.24079	0.74009	-0.21367

ESTIMATION FROM SAMPLES IN EXPONENTIAL & GAMMA MODELS

Table 7. Exponential distribution empirical sample characteristics of MMLEs of σ from left censored sample

		BIAS				VAR					MSE			
n	r	σ_M	σ_1	σ_2	σ_3	σ_M	σ_B	σ_1	σ_2	σ_3	σ_M	σ_1	σ_2	σ_3
3	1	0.00703	0.00231	0.00934	0.00887	0.32718	0.34211	0.34214	0.34219	0.34215	0.32723	0.34056	0.34870	0.34833
4	1	0.00041	0.00039	0.00349	0.00339	0.25367	0.25252	0.25253	0.25254	0.25253	0.25367	0.25233	0.25432	0.25426
	2	0.02048	0.00398	0.01613	0.01523	0.27658	0.26521	0.26526	0.26537	0.26528	0.27699	0.26317	0.27262	0.27366
5	1	0.00198	0.00011	0.00167	0.00164	0.20104	0.20098	0.20098	0.20098	0.20098	0.20104	0.20093	0.20165	0.20164
	2	0.00613	0.00080	0.00678	0.00655	0.20473	0.20523	0.20525	0.20527	0.20525	0.20477	0.20492	0.20810	0.20798
	3	0.01019	0.00520	0.02083	0.01963	0.22971	0.21966	0.21976	0.21984	0.21973	0.22981	0.21753	0.22952	0.22882
6	1	0.00057	0.00004	0.00122	0.00092	0.16817	0.16712	0.16712	0.16702	0.16712	0.16817	0.16711	0.16743	0.16743
	2	0.00737	0.00025	0.00351	0.00343	0.16318	0.16894	0.16895	0.16896	0.16895	0.16324	0.16887	0.17016	0.17013
	3	0.00210	0.00115	0.00946	0.00911	0.16966	0.17425	0.17427	0.17431	0.17427	0.16967	0.17387	0.17771	0.17755
	4	0.02233	0.00601	0.02419	0.02283	0.04620	0.18944	0.18949	0.18963	0.18951	0.20512	0.18725	0.19950	0.19878
7	1	0.00628	0.00024	0.00058	0.00063	0.14681	0.14309	0.14309	0.14309	0.14309	0.14685	0.14302	0.14326	0.14327
	2	0.00422	0.00008	0.00207	0.00203	0.14469	0.14401	0.14401	0.14401	0.14401	0.14471	0.14398	0.14461	0.14459
	3	0.00313	0.00037	0.00518	0.00504	0.13967	0.14645	0.14645	0.14646	0.14645	0.13968	0.14634	0.14801	0.14796
	4	0.00177	0.00143	0.01161	0.01117	0.15265	0.15239	0.15239	0.15245	0.15241	0.15265	0.15196	0.15614	0.15596
	5	0.03487	0.00384	0.02668	0.02522	0.18565	0.16786	0.16796	0.16804	0.16791	0.18687	0.16927	0.17784	0.17712

Table 7 (contd.). Exponential distribution empirical sample characteristics of MMLs of σ from left censored sample

n	r	BIAS				VAR					MSE			
		σ_M	σ_1	σ_2	σ_3	σ_M	σ_B	σ_1	σ_2	σ_3	σ_M	σ_1	σ_2	σ_3
8	1	0.00925	0.00000	0.00037	0.00037	0.12555	0.12514	0.12514	0.12514	0.12514	0.12563	0.12514	0.12523	0.12523
	2	0.01269	0.00002	0.00132	0.00129	0.12342	0.12564	0.12565	0.12565	0.12565	0.12358	0.12564	0.12626	0.12597
	3	0.00446	0.00004	0.00317	0.00309	0.12944	0.12693	0.12693	0.12694	0.12693	0.12946	0.12692	0.12775	0.12773
	4	0.01985	0.00019	0.00667	0.00642	0.13718	0.12979	0.12981	0.12982	0.12981	0.13757	0.12976	0.13161	0.13152
	5	0.01327	0.00188	0.01321	0.01279	0.14771	0.13612	0.13614	0.13619	0.13615	0.14789	0.13563	0.13998	0.13981
	6	0.02799	0.00755	0.02836	0.02702	0.15657	0.15162	0.15166	0.15182	0.15168	0.15736	0.14943	0.16134	0.16071
9	1	0.01059	0.00000	0.00026	0.00026	0.10773	0.11119	0.11119	0.11119	0.11119	0.10784	0.11119	0.11125	0.11125
	2	0.00398	0.00002	0.00088	0.00087	0.10740	0.11150	0.11149	0.11149	0.11149	0.10741	0.11149	0.11169	0.11169
	3	0.00990	0.00008	0.00207	0.00204	0.11348	0.11224	0.12224	0.11224	0.11224	0.11357	0.11222	0.11271	0.11270
	4	0.00332	0.00023	0.00414	0.00406	0.10735	0.11381	0.11382	0.11382	0.11381	0.10736	0.11376	0.11478	0.11476
	5	0.01077	0.00063	0.00781	0.00758	0.12049	0.11698	0.11698	0.11700	0.11698	0.12061	0.11684	0.11889	0.11883
	6	0.01098	0.00189	0.01469	0.01412	0.12623	0.12353	0.12355	0.12359	0.12568	0.12635	0.12309	0.12747	0.12945
	7	0.01943	0.00752	0.03000	0.02846	0.14248	0.13894	0.13897	0.13911	0.13898	0.14286	0.13695	0.14849	0.14785
10	1	0.00435	0.00000	0.00018	0.00018	0.10156	0.10006	0.10006	0.10006	0.10006	0.10158	0.10003	0.10007	0.10007
	2	0.00630	0.00001	0.00063	0.00062	0.09815	0.10025	0.10025	0.10025	0.10025	0.09819	0.10023	0.10035	0.10035
	3	0.00134	0.00004	0.00143	0.00142	0.09606	0.10070	0.10071	0.10072	0.10071	0.09607	0.10071	0.10100	0.10100
	4	0.01044	0.00011	0.00279	0.00275	0.10234	0.10164	0.10165	0.10166	0.10161	0.10245	0.10163	0.10224	0.10223
	5	0.00136	0.00028	0.00503	0.00493	0.10330	0.10343	0.10344	0.10448	0.10344	0.10398	0.10338	0.10556	0.10448
	6	0.01078	0.00072	0.00884	0.00857	0.11406	0.10680	0.10680	0.10681	0.10680	0.11418	0.10664	0.10879	0.10870
	7	0.00377	0.00205	0.01584	0.01523	0.11753	0.11348	0.11349	0.11353	0.11348	0.11755	0.11302	0.11741	0.11720
	8	0.02827	0.00780	0.03117	0.02961	0.13595	0.12873	0.12875	0.12888	0.12876	0.13674	0.12681	0.13801	0.13737

ESTIMATION FROM SAMPLES IN EXPONENTIAL & GAMMA MODELS

Table 8. Gamma (2) distribution empirical sample characteristic of reliability estimates in right censored samples

<i>n</i>	<i>r</i>	<i>R(x)</i>	BIAS		VAR		MSE	
			MLE	BLUE	MLE	BLUE	MLE	BLUE
3	1	0.9	0.05736	0.05572	0.01769	0.01745	0.02098	0.02056
4	1	0.9	0.03791	0.03717	0.00964	0.00957	0.01108	0.01095
	2	0.9	0.05728	0.05449	0.01736	0.01697	0.02064	0.01994
5	1	0.9	0.02948	0.02907	0.00660	0.00657	0.00747	0.00742
	2	0.9	0.03647	0.03517	0.00925	0.00912	0.01058	0.01035
	3	0.9	0.05637	0.05276	0.01764	0.01715	0.02082	0.01993
6	1	0.9	0.01921	0.01896	0.00434	0.00433	0.00471	0.00469
	2	0.9	0.02764	0.02690	0.00611	0.00605	0.00687	0.00678
	3	0.9	0.03724	0.03549	0.00921	0.00904	0.01060	0.01030
	4	0.9	0.05420	0.04995	0.01598	0.01541	0.01892	0.01790
7	1	0.9	0.01917	0.01900	0.00353	0.00352	0.00389	0.00388
	2	0.9	0.02165	0.02117	0.00406	0.00404	0.00453	0.00449
	3	0.9	0.02779	0.02677	0.00602	0.00595	0.00679	0.00666
	4	0.9	0.03534	0.03324	0.00907	0.00887	0.01031	0.00997
	5	0.9	0.05240	0.04762	0.01491	0.01430	0.01765	0.01657
8	1	0.9	0.01534	0.01522	0.00274	0.00274	0.00298	0.00297
	2	0.9	0.01915	0.01882	0.00328	0.00326	0.00365	0.00362
	3	0.9	0.02042	0.01976	0.00435	0.00431	0.00477	0.00470
	4	0.9	0.02414	0.02290	0.00538	0.00530	0.00597	0.00582
	5	0.9	0.03573	0.03330	0.00862	0.00841	0.00990	0.00951
	6	0.9	0.05261	0.04739	0.01540	0.01471	0.01816	0.01695
9	1	0.9	0.01248	0.01238	0.00238	0.00238	0.00253	0.00253
	2	0.9	0.01499	0.01475	0.00284	0.00283	0.00307	0.00305
	3	0.9	0.01760	0.01714	0.00331	0.00329	0.00362	0.00358
	4	0.9	0.01982	0.01900	0.00416	0.00411	0.00455	0.00447
	5	0.9	0.02368	0.02224	0.00506	0.00497	0.00562	0.00546
	6	0.9	0.03404	0.03136	0.00787	0.00764	0.00903	0.00862
	7	0.9	0.05323	0.04765	0.01601	0.01528	0.01884	0.01755
10	1	0.9	0.01131	0.01124	0.00193	0.00192	0.00205	0.00205
	2	0.9	0.01285	0.01267	0.00219	0.00219	0.00236	0.00235
	3	0.9	0.01493	0.01459	0.00281	0.00280	0.00303	0.00301
	4	0.9	0.01612	0.01553	0.00301	0.00299	0.00327	0.00323
	5	0.9	0.01874	0.01778	0.00380	0.00375	0.00415	0.00407
	6	0.9	0.02543	0.02379	0.00534	0.00523	0.00598	0.00579
	7	0.9	0.03207	0.02919	0.00772	0.00747	0.00875	0.00832
	8	0.9	0.05336	0.04741	0.01511	0.01435	0.01796	0.01660

Table 9. Gamma (2) distribution empirical sample characteristic of reliability estimates in left censored samples

n	r	$R(x)$	BIAS					VARIANCE					MSE				
			MLE	MET. I	MET. II	MET. III	BLUE	MLE	MET. I	MET. II	MET. III	BLUE	MLE	MET. I	MET. II	MET. III	BLUE
3	1	0.9	0.04216	0.04425	0.04173	0.04177	0.04360	0.01196	0.01220	0.01192	0.01192	0.01213	0.01374	0.01416	0.01366	0.01367	0.01403
4	1	0.9	0.02725	0.02787	0.02696	0.02696	0.02769	0.00656	0.00661	0.00654	0.00654	0.00660	0.00731	0.00739	0.00727	0.00727	0.00737
	2	0.9	0.02912	0.03225	0.02855	0.02861	0.03137	0.00715	0.00741	0.00711	0.00711	0.00733	0.00800	0.00845	0.00792	0.00793	0.00832
5	1	0.9	0.02414	0.02443	0.02396	0.02396	0.02097	0.00513	0.00515	0.00512	0.00512	0.00494	0.00571	0.00575	0.00569	0.00569	0.00538
	2	0.9	0.02355	0.02469	0.02315	0.02317	0.02078	0.00512	0.00520	0.00510	0.00510	0.00495	0.00567	0.00581	0.00563	0.00563	0.00538
	3	0.9	0.02171	0.02554	0.02111	0.02121	0.02063	0.00521	0.00547	0.00516	0.00517	0.00513	0.00568	0.00612	0.00561	0.00562	0.00556
6	1	0.9	0.01654	0.01669	0.01642	0.01642	0.01665	0.00364	0.00365	0.00364	0.00364	0.00365	0.00392	0.00393	0.00391	0.00391	0.00392
	2	0.9	0.01845	0.01898	0.01817	0.01817	0.01888	0.00368	0.00370	0.00366	0.00366	0.00370	0.00402	0.00406	0.00399	0.00399	0.00405
	3	0.9	0.01915	0.02167	0.01872	0.01875	0.02039	0.00406	0.00420	0.00404	0.00405	0.00414	0.00443	0.00467	0.00439	0.00440	0.00455
	4	0.9	0.02053	0.02494	0.01992	0.02006	0.02385	0.00467	0.00495	0.00463	0.00465	0.00489	0.00509	0.00557	0.00503	0.00505	0.00546
7	1	0.9	0.01762	0.01771	0.01753	0.01753	0.01769	0.00316	0.00316	0.00315	0.00315	0.00316	0.00347	0.00347	0.00346	0.00346	0.00347
	2	0.9	0.01591	0.01622	0.01571	0.01571	0.01615	0.00296	0.00297	0.00295	0.00295	0.00297	0.00321	0.00324	0.00320	0.00320	0.00323
	3	0.9	0.01612	0.01686	0.01578	0.01579	0.01674	0.00324	0.00327	0.00322	0.00322	0.00326	0.00350	0.00355	0.00347	0.00347	0.00355
	4	0.9	0.01644	0.01824	0.01598	0.01602	0.01794	0.00336	0.00345	0.00334	0.00334	0.00343	0.00363	0.00378	0.00359	0.00360	0.00376
	5	0.9	0.01657	0.02104	0.01596	0.01609	0.02009	0.00364	0.00387	0.00361	0.00362	0.00383	0.00392	0.00432	0.00387	0.00388	0.00423

ESTIMATION FROM SAMPLES IN EXPONENTIAL & GAMMA MODELS

Table 9 (contd.). Gamma (2) distribution empirical sample characteristic of reliability estimates in left censored samples

<i>n</i>	<i>r</i>	<i>R(x)</i>	BIAS					VARIANCE					MSE				
			MLE	MET. I	MET. II	MET. III	BLUE	MLE	MET. I	MET. II	MET. III	BLUE	MLE	MET. I	MET. II	MET. III	BLUE
8	1	0.9	0.01405	0.01411	0.01398	0.01398	0.01409	0.00239	0.00240	0.00239	0.00239	0.00240	0.00259	0.00259	0.00259	0.00259	0.00259
	2	0.9	0.01484	0.01503	0.01469	0.01469	0.01499	0.00254	0.00255	0.00254	0.00254	0.00255	0.00277	0.00278	0.00275	0.00275	0.00278
	3	0.9	0.01412	0.01455	0.01387	0.01387	0.01450	0.00273	0.00275	0.00272	0.00272	0.00275	0.00293	0.00296	0.00291	0.00291	0.00296
	4	0.9	0.01299	0.01395	0.01266	0.01267	0.01382	0.00270	0.00274	0.00269	0.00269	0.00273	0.00287	0.00293	0.00285	0.00285	0.00292
	5	0.9	0.01440	0.01647	0.01397	0.01399	0.01611	0.00279	0.00287	0.00277	0.00277	0.00286	0.00300	0.00314	0.00297	0.00297	0.00312
	6	0.9	0.01294	0.01795	0.01241	0.01253	0.01670	0.00313	0.00336	0.00311	0.00312	0.00331	0.00330	0.00369	0.00326	0.00327	0.00359
9	1	0.9	0.01136	0.01140	0.01130	0.01130	0.01138	0.00210	0.00210	0.00210	0.00210	0.00210	0.00223	0.00223	0.00223	0.00223	0.00223
	2	0.9	0.01205	0.01217	0.01194	0.01194	0.01215	0.00222	0.00222	0.00221	0.00221	0.00222	0.00236	0.00237	0.00235	0.00235	0.00237
	3	0.9	0.01201	0.01228	0.01182	0.01182	0.01224	0.00214	0.00215	0.00213	0.00213	0.00215	0.00228	0.00230	0.00227	0.00227	0.00230
	4	0.9	0.01122	0.01178	0.01095	0.01095	0.01170	0.00222	0.00224	0.00222	0.00222	0.00224	0.00235	0.00238	0.00234	0.00234	0.00238
	5	0.9	0.01235	0.01346	0.01200	0.01202	0.01332	0.00235	0.00239	0.00234	0.00234	0.00239	0.00250	0.00257	0.00248	0.00248	0.00256
	6	0.9	0.01326	0.01550	0.01281	0.01287	0.01515	0.00259	0.00268	0.00258	0.00258	0.00267	0.00277	0.00292	0.00274	0.00274	0.00290
	7	0.9	0.01067	0.01575	0.00034	0.01025	0.01453	0.00265	0.00285	0.00225	0.00263	0.00280	0.00276	0.00310	0.00225	0.00274	0.00302
10	1	0.9	0.01058	0.01060	0.01053	0.01053	0.01059	0.00172	0.00172	0.00172	0.00172	0.00172	0.00183	0.00183	0.00183	0.00183	0.00183
	2	0.9	0.01050	0.01058	0.01040	0.01040	0.01056	0.00172	0.00172	0.00172	0.00172	0.00172	0.00183	0.00183	0.00183	0.00183	0.00183
	3	0.9	0.01159	0.01177	0.01143	0.01143	0.01174	0.00195	0.00196	0.00195	0.00195	0.00196	0.00209	0.00210	0.00208	0.00208	0.00209
	4	0.9	0.01075	0.01114	0.01054	0.01055	0.01109	0.00188	0.00189	0.00187	0.00187	0.00189	0.00199	0.00201	0.00198	0.00198	0.00201
	5	0.9	0.01075	0.01143	0.01047	0.01048	0.01136	0.00199	0.00201	0.00198	0.00198	0.00201	0.00211	0.00214	0.00209	0.00209	0.00214
	6	0.9	0.01139	0.01262	0.01102	0.01104	0.01247	0.00208	0.00212	0.00207	0.00207	0.00211	0.00221	0.00228	0.00219	0.00219	0.00227
	7	0.9	0.01114	0.01352	0.01069	0.01076	0.01316	0.00227	0.00235	0.00225	0.00226	0.00234	0.00239	0.00253	0.00237	0.00237	0.00251
	8	0.9	0.01079	0.01613	0.01027	0.01042	0.01488	0.00263	0.00285	0.00261	0.00262	0.00280	0.00275	0.00311	0.00271	0.00273	0.00302

Table 10. Exponential distribution empirical sample characteristic of reliability estimates in left censored samples

<i>n</i>	<i>r</i>	<i>R(x)</i>	BIAS					VAR					MSE				
			MLE	MET. I	MET. II	MET. III	BLUE	MLE	MET. I	MET. II	MET. III	BLUE	MLE	MET. I	MET. II	MET. III	BLUE
3	1	0.9	0.03982	0.04103	0.03949	0.03955	0.03952	0.00954	0.00967	0.00951	0.00951	0.00935	0.01112	0.01135	0.01107	0.01108	0.01091
4	1	0.9	0.02706	0.02734	0.02689	0.02690	0.02848	0.00517	0.00519	0.00516	0.00516	0.00575	0.00590	0.00594	0.00588	0.00588	0.00656
	2	0.9	0.02632	0.02816	0.02586	0.02596	0.02821	0.00555	0.00568	0.00551	0.00552	0.00541	0.00624	0.00647	0.00618	0.00619	0.00621
5	1	0.9	0.02079	0.02088	0.02068	0.02069	0.02335	0.00363	0.00363	0.00362	0.00362	0.00357	0.00406	0.00407	0.00405	0.00405	0.00411
	2	0.9	0.02035	0.02091	0.02008	0.02010	0.02077	0.00347	0.00350	0.00346	0.00346	0.00383	0.00389	0.00394	0.00386	0.00387	0.00427
	3	0.9	0.02212	0.02453	0.02162	0.02175	0.02082	0.00408	0.00422	0.00406	0.00406	0.00383	0.00457	0.00483	0.00452	0.00454	0.00427
6	1	0.9	0.01727	0.01732	0.01721	0.01721	0.01784	0.00278	0.00278	0.00278	0.00278	0.00300	0.00308	0.00308	0.00308	0.00308	0.00332
	2	0.9	0.01742	0.01765	0.01724	0.01725	0.01769	0.00261	0.00262	0.00261	0.00261	0.00271	0.00292	0.00293	0.00290	0.00290	0.00302
	3	0.9	0.01739	0.01822	0.01707	0.01711	0.01875	0.00287	0.00290	0.00285	0.00286	0.00291	0.00317	0.00324	0.00314	0.00315	0.00326
	4	0.9	0.01736	0.02010	0.01685	0.01698	0.02028	0.00310	0.00323	0.00307	0.00308	0.00320	0.00340	0.00363	0.00336	0.00337	0.00361
7	1	0.9	0.01333	0.01335	0.01329	0.01329	0.01509	0.00196	0.00197	0.00196	0.00196	0.00226	0.00214	0.00214	0.00214	0.00214	0.00249
	2	0.9	0.01493	0.01505	0.01482	0.01483	0.01542	0.00213	0.00214	0.00213	0.00213	0.00241	0.00236	0.00236	0.00235	0.00235	0.00265
	3	0.9	0.01379	0.01415	0.01357	0.01358	0.01521	0.00206	0.00207	0.00205	0.00205	0.00202	0.00225	0.00227	0.00224	0.00224	0.00225
	4	0.9	0.01510	0.01613	0.01474	0.01479	0.01417	0.00221	0.00225	0.00220	0.00220	0.00224	0.00244	0.00251	0.00242	0.00242	0.00244
	5	0.9	0.01339	0.01525	0.01287	0.01302	0.01714	0.00242	0.00249	0.00240	0.00241	0.00240	0.00260	0.00273	0.00257	0.00258	0.00269

ESTIMATION FROM SAMPLES IN EXPONENTIAL & GAMMA MODELS

Table 10 (contd.). Exponential distribution empirical sample characteristic of reliability estimates in left censored samples

<i>n</i>	<i>r</i>	<i>R(x)</i>	BIAS					VAR					MSE				
			MLE	MET. I	MET. II	MET. III	BLUE	MLE	MET. I	MET. II	MET. III	BLUE	MLE	MET. I	MET. II	MET. III	BLUE
8	1	0.9	0.01136	0.01137	0.01133	0.01133	0.01145	0.00170	0.00170	0.00170	0.00170	0.00176	0.00183	0.00183	0.00183	0.00183	0.00190
	2	0.9	0.01422	0.01428	0.01414	0.01414	0.01219	0.00192	0.00192	0.00191	0.00191	0.00179	0.00212	0.00212	0.00211	0.00211	0.00194
	3	0.9	0.01206	0.01224	0.01190	0.01191	0.01291	0.00168	0.00169	0.00168	0.00168	0.00174	0.00183	0.00184	0.00182	0.00182	0.00190
	4	0.9	0.01117	0.01163	0.01092	0.01095	0.01395	0.00183	0.00184	0.00182	0.00182	0.00186	0.00195	0.00198	0.00194	0.00194	0.00205
	5	0.9	0.01224	0.01346	0.01187	0.01192	0.01382	0.00182	0.00185	0.00181	0.00181	0.00191	0.00197	0.00203	0.00195	0.00195	0.00210
	6	0.9	0.01170	0.01494	0.01121	0.01134	0.01492	0.00205	0.00216	0.00204	0.00204	0.00205	0.00219	0.00238	0.00216	0.00217	0.00228
9	1	0.9	0.01178	0.01179	0.01176	0.01176	0.01159	0.00144	0.00144	0.00144	0.00144	0.00160	0.00158	0.00158	0.00158	0.00158	0.00173
	2	0.9	0.00997	0.01001	0.00991	0.00991	0.01143	0.00134	0.00134	0.00134	0.00134	0.00162	0.00144	0.00144	0.00144	0.00144	0.00175
	3	0.9	0.01244	0.01255	0.01233	0.01233	0.01095	0.00157	0.00157	0.00156	0.00156	0.00155	0.00172	0.00173	0.00172	0.00172	0.00167
	4	0.9	0.01109	0.01136	0.01091	0.01091	0.01175	0.00149	0.00149	0.00148	0.00148	0.00155	0.00161	0.00162	0.00160	0.00160	0.00169
	5	0.9	0.01030	0.01090	0.01004	0.01006	0.01124	0.00151	0.00153	0.00151	0.00151	0.00141	0.00162	0.00165	0.00161	0.00161	0.00154
	6	0.9	0.01098	0.01233	0.01061	0.01067	0.01273	0.00165	0.00169	0.00164	0.00165	0.00165	0.00178	0.00184	0.00176	0.00176	0.00181
	7	0.9	0.01096	0.01436	0.01047	0.01063	0.01405	0.00172	0.00182	0.00171	0.00171	0.00194	0.00184	0.00203	0.00182	0.00183	0.00214
10	1	0.9	0.01025	0.01026	0.01024	0.01024	0.00879	0.00128	0.00128	0.00128	0.00128	0.00121	0.00139	0.00139	0.00139	0.00139	0.00129
	2	0.9	0.01030	0.01032	0.01026	0.01026	0.00821	0.00128	0.00128	0.00127	0.00127	0.00123	0.00138	0.00138	0.00138	0.00138	0.00130
	3	0.9	0.00974	0.00981	0.00966	0.00966	0.00946	0.00132	0.00132	0.00132	0.00132	0.00128	0.00142	0.00142	0.00141	0.00141	0.00137
	4	0.9	0.00902	0.00918	0.00888	0.00889	0.00971	0.00137	0.00137	0.00137	0.00137	0.00131	0.00145	0.00146	0.00144	0.00144	0.00141
	5	0.9	0.01030	0.01064	0.01009	0.01011	0.00963	0.00136	0.00136	0.00135	0.00135	0.00126	0.00146	0.00148	0.00145	0.00145	0.00135
	6	0.9	0.00979	0.01050	0.00952	0.00954	0.01009	0.00145	0.00147	0.00144	0.00145	0.00121	0.00155	0.00158	0.00154	0.00154	0.00131
	7	0.9	0.01081	0.01226	0.01042	0.01048	0.01022	0.00148	0.00151	0.00147	0.00147	0.00142	0.00159	0.00166	0.00158	0.00158	0.00152
	8	0.9	0.00903	0.01253	0.00856	0.00870	0.01249	0.00148	0.00157	0.00147	0.00148	0.00170	0.00157	0.00173	0.00155	0.00155	0.00186

Ridge Regression and Ill-Conditioning

Ghadban Khalaf
King Khalid University
Saudi Arabia

Mohamed Iguernane
King Khalid University
Saudi Arabia

Hoerl and Kennard (1970) suggested the ridge regression estimator as an alternative to the Ordinary Least Squares (OLS) estimator in the presence of multicollinearity. This article proposes new methods for estimating the ridge parameter in case of ordinary ridge regression. A simulation study evaluates the performance of the proposed estimators based on the Mean Squared Error (MSE) criterion and indicates that, under certain conditions, the proposed estimators perform well compared to the OLS estimator and another well-known estimator reviewed.

Keywords: Ordinary Least Squares, ill-condition, ridge regression, simulation.

Introduction

In regression problems the goal is usually to estimate the parameters in the general linear regression model

$$Y = X\beta + e \quad (1)$$

where Y is an $(n \times 1)$ response vector, X is an $(n \times p)$ matrix of n observations of p predictors. It is important to note that X is not a square matrix since the number of data values n is usually larger than the number of predictors of p . β is an $(p \times 1)$ vector of unknown regression parameters, and e is an $(n \times 1)$ vector of the random noise in the observed data vector Y , it is often assumed that they are distributed as Gaussian with $E(e) = 0$ and $Var(e) = \sigma^2$.

However, a method is needed to estimate the parameter vector β . The most common method is the least squared regression by finding the parameter values which minimize the sum of squared residuals, given by

Dr. Khalaf is an Associate Professor in the Department of Mathematics. Email him at albadran50@yahoo.com. Dr. Iguernane is an Assistant Professor in the Department of Mathematics. Email him at Mohamed.iguernane@gmail.com.

$$SSR = \sum |Y - X\beta|^2 \quad (2)$$

The solution turns out to be a matrix equation, defined by

$$\hat{\beta} = (X'X)^{-1} X'Y \quad (3)$$

where X' is the transpose of the matrix X and the exponent (-1) indicates the matrix inverse of the given quantity.

It is expected that the true parameters will provide the most likely result, so the least squares solutions, by minimizing the sum of squared residuals, gives the maximum likelihood values of the parameters vector β . It is known from the Gauss-Markov theorem that the least squares estimate results the best linear unbiased estimator of the parameters; thus, this is one reason why least squares method is very popular. The estimates of the least squares are unbiased (i.e., the expected values of the parameters are the true values), and of all the unbiased estimators, it gives the least variance.

However, there are cases for which the best linear unbiased estimator is not necessarily the best estimator. One pertinent case occurs when the two (or more) of the predictor variables are very strongly correlated. In other words, when terms are correlated and the columns of the design matrix X have an approximate linear dependence, the matrix $(X'X)^{-1}$ becomes close to singular. As a result, the least squares estimate, given by (3), becomes highly sensitive to random errors in the observed response Y , producing a large variance. To solve this problem, one approach is to use an estimator which is no longer unbiased, but has considerably less variance than the least squares estimator.

Ridge Regression and Multicollinearity

Ridge Regression is a technique for analyzing multiple regression data that suffer from multicollinearity. When multicollinearity occurs, least squares estimates are unbiased but their variances are large so they may be far from the true value, deflate the partial t -test for the regression coefficients give false non-significant p -values and degrade the predictability of the model. Thus, by adding a degree of bias to the regression estimates, ridge regression reduces the standard errors and the matrix needed to invert no longer has a determinant near zero; therefore, the solution does not lead to uncomfortably large variance in the estimated parameters. Now, given a response vector Y and a predictor matrix X , the ridge regression coefficients are given by

$$\hat{\beta}(k) = (X'X + kI)^{-1} X'Y \quad (4)$$

where k is the ridge parameters and I is the identity matrix. When $k = 0$, the linear regression estimate is given by (3), and when $k = 1$, $\hat{\beta}(k) = 0$, finally, for k in between, two ideas are balanced: fitting a linear model of Y on X 's and shrinking the coefficients. Small positive values of k improve the conditioning of the problem and reduce the variance of the estimates. While biased, the reduced variance of ridge estimates often result in a smaller MSE when compared to least squares estimates.

The amount of shrinkage is controlled by k , the ridge parameter that multiplies the ridge penalty. Large k means more shrinkage, thus, different coefficient estimates are obtained for different values of k . In fact, choosing an appropriate value of k is important and also difficult, but it can be shown that there exists a value of k for which the MSE (the variance plus the bias squared) of the ridge estimator is less than that of the least squares estimator. As a result, under the condition of multicollinearity, a huge price is paid for the unbiasedness property that is achieved by using the OLS estimator.

Choosing the Ridge Parameter k

One of the main obstacles in using ridge regression is in choosing an appropriate value of k . For selecting the best ridge estimator, several criteria have been proposed in the literature (see for example; Hoerl & Kennard, 1970; Hoerl et al., 1975; Hoerl & Kennard, 1976; Lawless & Wang, 1976; Gibbons, 1981; Saleh & Kibria, 1993; Troskie & Chalton, 1996; Kibria, 2003; Khalaf & Shukur, 2005; Dorugade & Kashid, 2010; and Khalaf, 2013). Next, some formulas for determining the value of k to be used in (4) are discussed.

Hoerl and Kennard (1970) suggested using a graphic which they called the ridge trace. This plot shows the ridge regression coefficients as a function of k . When viewing the ridge trace, the value of k is chosen at which the regression coefficients have reasonable magnitude, sign and stability, while the MSE is not grossly inflated. In fact, letting β_{\max} denote the maximum of the B_i , Hoerl and Kennard (1970) showed that choosing

$$\hat{k} = \frac{\hat{\sigma}^2}{\hat{\beta}_{\max}^2} \quad (5)$$

implies that $MSE(\hat{\beta}(k)) < MSE(\hat{\beta}) = \hat{\sigma}^2 \sum_{i=1}^p t_i^{-1}$, where $\hat{\sigma}^2$ is the usual estimator of σ^2 , defined by $\hat{\sigma}^2 = \frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{n - p - 1}$. The estimator, given by (5), will be denoted by HK.

Hoerl, Kennard and Baldwin (1975) argued that a reasonable choice of k is

$$k = \frac{p\sigma^2}{\beta'\beta} \quad (6)$$

if these quantities were known. They suggested using

$$\hat{k} = \frac{p\hat{\sigma}^2}{\hat{\beta}'\hat{\beta}} \quad (7)$$

as an estimate of k in (6). This ridge estimator will be denoted by HKB.

Hoerl and Kennard (1976) proposed an iterative method for selecting k . This method is based on the formula given by (7). To obtain the first value of k , they used the least squares coefficients. This produces a value of k . Using this new k , a new set of coefficients is found, and so on. In fact, this procedure does not necessarily converge.

Lawless and Wang (1976) concluded that the ridge estimators using (5) and (7) performed very well indeed and that they were substantially better than any of the other estimators included in their study. Gibbons (1981) conducted a simulation study to compare 10 promising algorithms for selecting k . She found too that the estimators using the ridge estimator given by (7) performed well. In the light of these remarks, which indicate the satisfactory performance and the potential for improvement of the estimators HK and HKB, new methods are proposed to determine ridge parameter in case of ordinary ridge regression for the ridge parameter k as

$$\begin{aligned} 1) \quad K I_a &= \text{The Arithmetic Mean of (HK, HKB)} \\ &= \frac{\hat{\sigma}^2}{2} \left(\frac{1}{\hat{\beta}_{\max}^2} + \frac{p}{\hat{\beta}'\hat{\beta}} \right) \end{aligned} \quad (8)$$

$$\begin{aligned}
 2) \quad KI_h &= \text{The Harmonic Mean of (HK, HKB)} \\
 &= \frac{2}{\frac{1}{HK} + \frac{1}{HKB}} = \frac{2\hat{\sigma}^2}{\hat{\beta}_{\max}^2 + \frac{\hat{\beta}'\hat{\beta}}{p}} \quad (9)
 \end{aligned}$$

$$\begin{aligned}
 3) \quad KI_g &= \text{The Geometric Mean of (HK, HKB)} \\
 &= \sqrt{HK \cdot HKB} = \hat{\sigma}^2 \sqrt{\frac{p}{\hat{\beta}_{\max}^2 \cdot \hat{\beta}'\hat{\beta}}} \quad (10)
 \end{aligned}$$

$$4) \quad KI_s = \begin{cases} \text{The sum of (HK, HKB),} & \text{if } HK + HKB < 1 \\ \text{The sum of (HK, HKB) / 2,} & \text{if } HK + HKB \geq 1 \end{cases} \quad (11)$$

If the resulting $HK+HKB$, given by (11), is less than one, then it is used as an estimator for the ridge parameter k . However, if the resulting $HK+HKB$ is greater than or equal to one then the new value of the ridge parameters equal to the value of $(HK+HKB)$ divided by two.

Simulation Study

A simulation study was conducted in order to draw conclusions about the performance of the proposed estimators relative to HK, HKB and the OLS estimator. To achieve different degrees of collinearity, following Kibria (2003), the independent variables were generated by using the following equation

$$x_{ij} = (1 - \rho^2)^{\frac{1}{2}} z_{ij} + \rho z_{ip}, \quad i = 1, 2, \dots, n \quad j = 1, 2, \dots, p \quad (12)$$

where z_{ij} are independent standard normal distribution, p is the number of the explanatory variables and ρ is specified so that the correlation between any two independent variables is given by ρ^2 . Four different sets of correlation were considered according to the value of $\rho = 0.7, 0.9, 0.95$ and 0.99 .

The other factors varied were sample size (n) and the number of regressors (p). Models consisting of 15, 25, 50 and 100 observations and with 5 and 9 explanatory variables were generated.

The criterion proposed for measuring the goodness of an estimator is the MSE using the following formula

$$MSE = \frac{1}{2000} \sum_{i=1}^p (\hat{\beta}_i - \beta)' (\hat{\beta}_i - \beta), \quad (13)$$

where $\hat{\beta}_i$ is the estimator of β obtained from the OLS estimator or from the ridge estimator for different estimated value of k considered for comparison reasons and, finally, 2000 is the number of replications used in the simulation. In this study the error was forced to have variances equal to 0.5 and 1.

Simulation results show that increasing the number of regressors leads to a higher estimated MSE, while increasing the sample size leads to a lower estimated MSE (see Khalaf & Shukur, 2005; Alkhamisi & Shukur, 2008; Khalaf, 2011).

Results

Tables 1 and 2 present the output of the simulation concerning properties of the different methods that used to choose the ridge parameter k .

Results show that the estimated MSE is affected by all factors that were varied. It is also noted that the higher the degree of correlation the higher estimated MSE, but this increase is much greater for the OLS than the ridge regression estimator. The sample size and the number of explanatory variables having a different impact of the estimators.

In Tables 1 and 2 when $\rho = 0.7$ and n is large, note that the estimated MSE decreases substantially and the performance of KI_s is much better than the other ridge estimators from the MSE point of view. Finally, the OLS estimator is defeated by all of estimators.

Conclusion

Ridge regression is one of the more popular estimation procedures for addressing issues of multicollinearity. The procedures discussed herein fall into the category of biased estimation techniques. They are based on this notion: though the OLS gives unbiased estimates and indeed enjoy the minimum variance of all linear unbiased estimators, there is no upper bound on the variance of the estimators and the presence of multicollinearity may produce large variance. Biased estimation is used to attain a substantial reduction in variance with an accompanied increase in stability of the regression coefficients. The coefficients become biased, but the reduction in variance is of greater magnitude than the bias induced in the estimators.

New methods were proposed for estimating the ridge parameters in the presence of multicollinearity. The performance of the proposed ridge parameter was evaluated through the simulation, for different combinations of correlation between predictors (ρ), the number of explanatory variables (p), sample size (n) and variance of the error variable (σ^2). The evaluation of the estimators was done by comparing the MSE of the OLS estimator with the proposed estimators and the other estimators reviewed in this study. Finally, it was found that the performance of the proposed estimators is satisfactory over the others and KI_s has the least MSE.

Table 1. Estimated MSE when $p = 5$

ρ	σ^2	n	OLS	HK	HKB	KI_a	KI_g	KI_h	KI_s
0.7	0.05	15	4.53	2.75	1.92	2.23	2.34	2.44	1.68
		25	2.13	1.63	1.19	1.37	1.43	1.48	1.14
		50	0.95	0.84	0.67	0.74	0.77	0.79	0.68
		100	0.42	0.40	0.36	0.38	0.38	0.39	0.36
	1	15	1.16	0.95	0.73	0.82	0.85	0.88	0.73
		25	0.54	0.50	0.42	0.45	0.47	0.48	0.43
		50	0.23	0.22	0.20	0.21	0.21	0.22	0.21
		100	0.1025	0.1013	0.0981	0.0995	0.1009	0.1001	0.0991
0.9	0.05	15	14.06	6.13	4.44	4.95	5.25	5.56	3.08
		25	6.44	3.62	2.40	2.86	3.01	3.17	1.90
		50	2.83	2.04	1.35	1.62	1.72	1.81	1.21
		100	1.27	1.08	0.79	0.90	0.95	0.99	0.78
	1	15	3.510	2.209	1.475	1.752	1.851	1.948	1.321
		25	1.633	1.293	0.908	1.064	1.121	1.174	0.892
		50	0.7063	0.6391	0.5068	0.5643	0.5867	0.6061	0.5224
		100	0.3279	0.3130	0.2744	0.2921	0.2990	0.3046	0.2841
0.95	0.05	15	28.00	10.00	7.00	8.43	9.00	10.00	4.00
		25	13.97	6.09	4.31	4.87	5.17	5.48	2.89
		50	5.91	3.41	2.21	2.67	2.82	2.97	1.72
		100	2.78	2.02	1.33	1.61	1.70	1.79	1.18
	1	15	6.91	3.53	2.34	2.76	2.92	3.09	1.82
		25	3.38	2.26	1.49	1.79	1.89	1.99	1.31
		50	1.46	1.19	0.84	0.99	1.04	1.09	0.83
		100	0.68	0.62	0.49	0.54	0.57	0.59	0.50
0.99	0.05	15	156.00	49.00	38.00	36.00	41.00	47.00	20.00
		25	73.00	23.00	18.26	18.08	20.00	22.00	10.00
		50	32.00	12.00	8.00	9.00	10.00	11.00	5.00
		100	15.00	6.76	4.71	5.35	5.70	6.06	3.08
	1	15	39.00	14.17	10.35	10.73	11.90	13.17	6.25
		25	18.00	7.37	5.33	5.82	6.26	6.72	3.35
		50	8.16	4.17	2.79	3.29	3.48	3.67	2.01
		100	3.78	2.43	1.56	1.89	2.01	2.11	1.29

RIDGE REGRESSION AND ILL-CONDITIONING

Table 2. Estimated MSE when $p = 9$

ρ	σ^2	n	<i>OLS</i>	<i>HK</i>	<i>HKB</i>	<i>KI_a</i>	<i>KI_g</i>	<i>KI_h</i>	<i>KI_s</i>
0.7	0.05	15	16.29	8.70	4.76	5.98	6.65	7.35	4.17
		25	5.38	4.18	2.50	3.09	3.41	3.71	2.40
		50	2.03	1.86	1.35	1.55	1.67	1.76	1.38
		100	0.89	0.86	0.73	0.78	0.82	0.84	0.74
	1	15	4.04	2.93	1.7483	2.13	2.35	2.58	1.7482
		25	1.33	1.23	0.93	1.04	1.11	1.17	0.96
		50	0.48	0.47	0.42	0.44	0.45	0.46	0.43
		100	0.216	0.214	0.203	0.208	0.211	0.213	0.206
0.9	0.05	15	48.00	19.00	11.00	13.00	15.00	16.00	8.00
		25	16.63	9.46	5.07	6.52	7.23	7.95	4.19
		50	6.28	4.83	2.72	3.46	3.87	4.25	2.51
		100	2.72	2.40	1.55	1.87	2.06	2.22	1.53
	1	15	12.36	6.74	3.52	4.50	5.07	5.66	3.14
		25	4.00	3.19	1.85	2.31	2.59	2.84	1.82
		50	1.52	1.40	1.00	1.16	1.25	1.33	1.03
		100	0.70	0.67	0.56	0.61	0.64	0.66	0.58
0.95	0.05	15	105.00	38.00	21.00	24.00	28.00	33.00	14.00
		25	34.18	16.43	8.87	11.20	12.50	13.85	6.79
		50	12.62	8.01	4.21	5.49	6.14	6.78	3.55
		100	5.78	4.47	2.50	3.19	3.58	3.94	2.29
	1	15	25.25	11.93	6.38	7.98	9.01	10.08	5.15
		25	8.65	5.84	3.08	4.00	4.50	4.98	2.80
		50	3.10	2.63	1.60	1.97	2.20	2.39	1.59
		100	1.36	1.26	0.89	1.04	1.13	1.19	0.92
0.99	0.05	15	563.00	197.00	111.00	122.00	146.00	173.00	73.00
		25	185.00	70.00	40.00	46.00	53.00	61.00	26.00
		50	71.00	30.00	17.00	20.00	23.00	26.00	12.00
		100	31.00	15.00	8.00	10.00	12.00	13.00	6.00
	1	15	140.00	48.00	28.00	31.00	36.00	42.00	18.00
		25	48.95	21.11	11.95	14.40	16.24	18.15	8.62
		50	18.09	10.11	5.29	6.86	7.67	8.48	4.30
		100	7.96	5.49	2.93	3.82	4.32	4.80	2.62

References

- Alkhamisi, M., & Shukur, G. (2008). Developing ridge parameters for SUR model. *Communications in Statistics, Theory and Methods*, 37, 544-564.
- Dorugade, A. V., & Kashid, D. N. (2010). Alternative method for choosing ridge parameter for regression. *International Journal of Applied Mathematical Sciences*, 4(9), 447-456.
- Gibbons, D. G. (1981). A simulation study of some ridge estimators. *Journal of the American Statistical Association*, 76(373), 131-139.
- Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for non-orthogonal Problems. *Technometrics*, 12, 55-67.
- Hoerl, A. E., & Kennard, R. W. (1976). Ridge Regression: Iterative Estimation of the biasing Parameter. *Communications in Statistics*, 5, 77-88.
- Hoerl, A. E., Kennard, R. W., & Baldwin, K. F. (1975). Ridge Regression: some Simulation. *Communications in Statistics - Theory and Methods*, 4, 105-124.
- Khalaf, G. (2011). Ridge Regression: An Evaluation to some New Modifications. *International Journal of Statistics and Analysis*, 1(4), 325-342.
- Khalaf, G. (2013). A Comparison Between Biased and Unbiased Estimators. *Journal of Modern Applied Statistical Methods*, 12(2), 293-303. Retrieved from <http://digitalcommons.wayne.edu/jmasm/vol12/iss2/17/>
- Khalaf, G., & Shukur, G. (2005). Choosing Ridge Parameters for Regression Problems. *Communication in Statistics – Theory and Methods*, 34, 1177-1182.
- Kibria, B. M. G. (2003). Performance of some New Ridge Regression Estimators. *Communication in Statistics - Theory and Methods*, 32, 419-435.
- Lawless, J. P., & Wang, P. A. (1976). A Simulation Study of Ridge and Other Regression Estimators. *Communications in Statistics*, 5, 307-323.
- Saleh, A. K., & Kibria, B. M. (1993). Performances of some new preliminary test ridge regression estimators and their properties. *Communications in Statistics - Theory and Methods*, 22, 2747-2764.
- Troskie, C. G., & Chalton, D. O. (1996). A Bayesian estimate for the constants in ridge regression. *South African Statistical Journal*, 30, 119-137.

Comparison of Individual and Moving Range Chart Combinations to Individual Charts in Terms of ARL after Designing for a Common “All OK” ARL

Dewi Rahardja

U.S. Department of Defense
Indianapolis, Indiana

In some process monitoring situations, consecutive measurements are spaced widely apart in time, making monitoring process aim and spread difficult. This study uses three cases to compare the effectiveness of two such monitoring schemes, i.e., the X chart alone (X-only chart) and the Individuals and Moving Range Chart Combination (X/MR charts), in terms of Average Run Length (ARL) after designing for a common “all OK” (in-control) ARL. The study finds that X chart alone is sufficient (and hence, recommended) in detecting changes in all the 3 cases: changes in the process mean, changes in the process standard deviation, and changes in both process mean and standard deviation.

Keywords Individual chart, X-chart alone, moving range chart, X/MR chart, ARL, Average Run Length, “all OK” ARL

Introduction

In some process monitoring situations, consecutive measurements are spaced widely apart in time. For example, an engineering process may allow only one measurement per day. In some cases, a series of individual items are produced in such a way that no natural subgrouping is possible (Crowder, 1987a). When this happens, exactly how to monitor process aim and spread is not completely obvious. One sensible possibility is to simply plot individual observations on their own chart (X-only chart). Another possibility is to plot a combination of a chart for individual measurements and a moving range chart based on two consecutive observations. Duncan (1974) outlines such a procedure.

The purpose of this study is to compare the effectiveness of these two monitoring schemes, i.e., the X chart alone (X-only chart) and the Individuals and

Contact Dewi Rahardja at Dewi.G.Rahardja.civ@mail.mil or rahardja@gmail.com.

Moving Range Chart Combination (X/MR chars), in terms of Average Run Length (ARL) after designing for a common “all OK” (in-control) ARL.

The run length of any process monitoring procedure is the number of sampling periods before an out-of-control signal is given. An out-of-control signal suggests that some change in the process has occurred and that action should be taken to find and correct any assignable causes. The average run length (ARL) is often used to describe the likely performance of a control procedure. A large ARL is desired when the process is stable or in control, and a small ARL otherwise (Crowder, 1987a).

Comparison of monitoring schemes will be made under three sets of circumstances. The first case is where the process mean changes from its standard value, the second case is where the process variability changes, and third case is where both process mean and process variability change from standard values. In each of these three cases, a small ARL is desired, since it will indicate quick detection of the out-of-control situation.

Literature Review

Vardeman and Jobe (1999) discussed the charting of individuals and moving ranges and some other process monitoring techniques that improve on Shewhart charts in situations where it is important to quickly detect small process changes. That is, they also considered EWMA and CUSUM process monitoring schemes. Four types of process monitoring schemes were originally considered in the present study: The X chart alone, Individuals and Moving Range Chart Combinations, EWMA and CUSUM process monitoring schemes. However, because EWMA and CUSUM schemes are known to be better than an X chart alone for detecting small process changes, no further analysis is needed (for EWMA and CUSUM) if the X chart alone is better than Individual and Moving Range Chart Combinations.

Crowder (1987a, 1987b) discussed the Computation of ARLs for Combined Individual Measurement and Moving Range Charts. Numerical procedures and a control chart design strategy are presented. ARLs are given for various choices of the control limits and shifts in the level of the process mean and standard deviation. Also, a Fortran computer program was presented that allows inputting control limits for combined individual measurement and moving range charts and then returns the approximate average run length (ARL) for the normal case with standard deviation 1 and various shifts in the process mean.

Roes, et al. (1993) discussed several options in designing a Shewhart-type control chart for Individual Observations. A number of possible estimators of the

X VS. X/MR CHARTS

standard deviation were considered and a two-stage procedure is suggested for retrospective testing. It was argued that adding a Moving Range Chart has no real added value and, therefore, was ill-advised.

Rigdon, et al. (1994) discussed design strategies for Individuals and Moving Range Control Charts. These authors argue that an X chart alone is nearly as efficient as the combined X/MR chart for detecting changes in the process variability. For the same in-control ARL, the X-only chart is more effective than the X/MR chart combination with moving ranges of $k = 2, 3, 4$ for detecting shifts in the process mean, while the two schemes are about equally effective in detecting changes in the process variability. The size (k) of the window for the moving range has little effect on the ARL for changes in the process variability.

Radson, et al. (1995) considered the possibility of a shift from the in-control standard deviation, σ_0 , to a standard deviation level of σ_1 , where $\sigma_1 = k \cdot \sigma_0$. In their study, k ranged from 0.1 to 3.0 in increment of 0.1. They demonstrated that the moving range can be used to detect variance reduction. This was best achieved by constructing limits based on the true underlying distribution for the moving range.

Adke and Hong (1997) discussed the X chart and the Moving Range chart for the normally distributed observations when there is a shift in the process variance. They concluded that a Moving Range chart does provide useful information.

Amin (1998) showed that there is no disadvantage in using an X/MR procedure. The discussion is limited to normality assumption.

Marks and Krehbiel (2009) evaluated the Individual Chart and X/MR Chart Combinations for just the first 2 cases mentioned above: mean change only and variability change only, from standard values. But they did not consider the third case, where both mean and variability changes occur, from standard values.

Methods

Several “all OK” (in-control) ARLs are chosen and control limits set to produce them (The “all OK” ARL is the ARL when the process mean and process standard deviation are on-target, i.e., the process mean and process standard deviation are equal to standard values). Then, supposing the mean and/or standard deviation are off-target, new ARLs for both the X/MR Chart Combinations and the X-only Chart are computed and compared (a smaller off-target ARL is preferred since the corresponding charting method then gives quicker detection of non-standard conditions).

X-only Chart Monitoring Scheme

First, an “all OK” (in-control) ARL is chosen. Then, using $\mu_0 = 0$ and $\sigma_0 = 1$ as the standard conditions, control limits for the X-only Chart are

$$\begin{aligned} UCL_{X-only} &= \mu_0 + L\sigma_0 = L \\ LCL_{X-only} &= \mu_0 - L\sigma_0 = -L \end{aligned}$$

where the given “all OK” (in-control) ARL is simply $1/r$ for

$$\begin{aligned} r &= P[1^{st} \text{ point plots outside the control limits}] \\ r &= P[Z > L] + P[Z < -L] \end{aligned}$$

which can be calculated from the standard normal distribution table.

After control limits are set, the off-target ARL of the X-only Chart for any other μ and σ combination, i.e., non-standard conditions, can be computed as follows.

Let

$$\begin{aligned} \mu_{new} &= \mu_0 + \delta = \mu_0 + D\sigma_{new} = \mu_0 + DT\sigma_0 \\ \sigma_{new} &= T\sigma_0 \end{aligned} \tag{1}$$

where $D = \text{constant}$ (i.e., a shift in the process mean) and $T = \text{constant}$ (i.e., a change in the process standard deviation).

Then,

$$z_1 = \frac{UCL_{X-only} - \mu_{new}}{\frac{\sigma_{new}}{\sqrt{n}}} = \frac{L - \mu_{new}}{\sigma_{new}} = a$$

and

$$z_2 = \frac{LCL_{X-only} - \mu_{new}}{\frac{\sigma_{new}}{\sqrt{n}}} = \frac{-L - \mu_{new}}{\sigma_{new}} = b$$

X VS. X/MR CHARTS

for $n = 1$, since the natural sample size is one. Finally,

$$r_{new} = P[Z > a] + P[Z < b]$$

and the new ARL, i.e., the off-target ARL is $1/r_{new}$.

X/MR Chart Combinations Monitoring Scheme

Crowder (1987a, 1987b) noted that ARLs of X/MR chart combinations can be obtained by solving certain integral equations as follows. ARLs for control schemes are typically evaluated under the model $y_t = \mu + \varepsilon_t$, $t = 1, 2, 3, \dots$, where ε_t 's are independent $N(0, \sigma^2)$ random variables and y is the observation made at time t . Suppose we wish to control the process mean and standard deviation at nominal levels $\mu_0 = 0$ and $\sigma_0 = 1$. The X/MR procedure is to plot on separate charts the individual observations y_t and the successive moving ranges $r_t = |y_t - y_{t-1}|$. The process is deemed out-of-control at observation one if $|y_t| > M$ and out-of-control at time $t > 1$ if either $|y_t| > M$ or $r_t > R$, where M and R are specified positive constants.

Let $L(u)$ be the mean additional time until an out-of-control signal, given that the most recent observation is u . If the next observation, y , is more than R units from u or is larger in magnitude than M , an out-of-control signal is given. Otherwise, the run continues with y as the most recent observation. Thus, supposing that observations are normal with mean μ and variance σ^2 ,

$$\begin{aligned} L(u) &= 1 \cdot P(|y - u| > R \text{ or } |y| > M) + \int_{(x: x-u \leq R \text{ and } |x| \leq M)} [1 + L(x)] \cdot f(x) dx \\ &= 1 + \int_{\max\{-M, u-R\}}^{\min\{M, u+R\}} L(x) f(x) dx \end{aligned} \quad (2)$$

where $f(x)$ is the $N(0, \sigma^2)$ density. Now letting T be the run length associated with the procedure and conditioning on y_1 ,

$$\begin{aligned}
ARL &= E(T) \\
&= E[E(T|y_1)] \\
&= E[E(T|y_1) \text{ where } |y_1| > M] + E[E(T|y_1) \text{ where } |y_1| \leq M]. \quad (3) \\
&= 1 \cdot P(|y_1| > M) + \int_{-M}^M [1 + L(x) \cdot f(x) dx] \\
&= 1 + \int_{-M}^M L(x) \cdot f(x) dx
\end{aligned}$$

Note that $L(x)$ in (3) is unknown. Thus, to approximate the ARL, values of $L(x)$ on the interval $(-M, M)$ are first approximated using (2). The solution to the integral equation (2) can be obtained by replacing the integral equation with a system of linear algebraic equations and solving them numerically using trapezoidal quadrature. The ARL can then be approximated via (3), again using trapezoidal quadrature and the approximations of $L(x)$ at each of the subinterval points. Crowder's published Fortran program (1987b) will compute this ARL.

Vardeman and Jobe (1999) identified several M/R combinations giving each of several choices of "all OK" (in-control) ARLs as shown in Table 1. These M/R combinations will be compared to the X-only possibility.

Table 1. Several M/R combinations for various choices of "all OK" ARLs

"all OK" ARL	Smallest <i>M</i> possible				Smallest <i>R</i> possible
50	M = 2.33 (R = 4.5+)	M = 2.40 R = 3.66	M = 2.55 R = 3.41	M = 2.80 R = 3.28	(M = 3.3+) R = 3.24
100	M = 2.58 (R = 5.0+)	M = 2.65 R = 4.04	M = 2.80 R = 3.77	M = 3.00 R = 3.67	(M = 3.5+) R = 3.60
250	M = 2.88 (R = 5.5+)	M = 2.95 R = 4.47	M = 3.10 R = 4.22	M = 3.30 R = 4.11	(M = 3.80+) R = 4.05
370	M = 3.00 (R = 6.0+)	M = 3.10 R = 4.40	M = 3.20 R = 4.40	M = 3.40 R = 4.29	(M = 3.8+) R = 4.23
500	M = 3.09 (R = 6.0+)	M = 3.30 R = 4.53	M = 3.30 R = 4.53	M = 3.50 R = 4.42	(M = 4.0+) R = 4.36
750	M = 3.21 (R = 6.0+)	M = 3.45 R = 4.88	M = 3.45 R = 4.66	M = 3.60 R = 4.59	(M = 4.0+) R = 4.55
1000	M = 3.29 (R = 6.5+)	M = 3.40 R = 4.96	M = 3.50 R = 4.82	M = 3.65 R = 4.72	(M = 4.0+) R = 4.65

Source: Vardeman & Jobe (1999)

X VS. X/MR CHARTS

Because Crowder's published program requires M and R values for inputs and outputs $\sigma = 1$ ARLs for various means, the following procedures are necessary to use the program to evaluate the ARLs we desire.

As before, without loss of generalities, standard values $\mu_0 = 0$ and $\sigma_0 = 1$ are used. From expression (1), then with

$$\begin{aligned}UCL_X &= \mu_0 + M\sigma_0 = M \\LCL_X &= \mu_0 - M\sigma_0 = -M\end{aligned}$$

and

$$UCL_{MR} = R \cdot \sigma_0 = R ,$$

one must use limits

$$M_{new} = \frac{UCL_X - LCL_X}{2 \cdot \sigma_{new}} = \frac{M}{T}$$

and

$$R_{new} = \frac{UCL_{MR}}{\sigma_{new}} = \frac{R}{T}$$

must be used with Crowder's program (1987b).

Notice that for M large (The "Smallest Possible R " values in Table 1), the ARLs can essentially be considered ARLs of the Moving Chart alone (Crowder, 1987a). Similarly, for R large (The "Smallest Possible M " values in Table 1), the ARLs can essentially be considered ARLs of the Individual Chart alone. Therefore, the smallest possible M 's from Table 1 are the same as the L 's obtained for the Individual Chart alone.

Results

A variety of "all OK" (in control) ARLs, namely 50, 100, 250, 370, 500, 750 and 1000 from Vardeman and Jobe (1999) are selected from this study.

Comparisons of the monitoring schemes under the sets of circumstances mentioned in the Introduction are made in Tables 2-4 and Figures 1-3. Because the results for the different “all OK” ARLs all turned out to be similar, only Tables and Figures for the ARL = 370 case are presented here. In Tables 2-4, the smallest ARLs for a given set of parameters are bold, indented, and underlined. Table 2 and the corresponding Figure 1 show the case where only the process mean changes. Table 3 and the corresponding Figure 2 show the case where only the process standard deviation changes. Finally, Table 4 and Figure 3 show the case where both process mean and process standard deviation change from standard values.

Mean Changes Only (Table 2 and Figure 1)

As shown in Table 2, the smallest ARLs (bold, indented, and underlined) fall either in the X-only Chart region or in X/MR Chart Combinations region with R large (Smallest Possible M). As noted in ‘Methods’ above, X/MR Chart Combinations with R large (Smallest Possible M) can essentially be considered the X-only Charts. Therefore, the first 2 “ARL-Columns” (from left) in all the Tables 2-4 are in fact equivalent.

Table 2. Mean Changes – ARL = 370

D	X-only	X/MR Chart Combinations				
		M = 3.00 R = 6.0+	M = 3.10 R = 4.57	M = 3.20 R = 4.40	M = 3.40 R = 4.29	M = 3.8+ R = 4.23
0.00	370.0	370.0	367.8	370.9	363.0	374.8
0.25	<u>277.8</u>	280.9	299.7	318.4	336.5	369.6
0.50	156.3	<u>155.1</u>	182.9	212.5	265.1	349.2
0.75	81.3	<u>81.2</u>	100.5	123.3	177.3	303.6
1.00	<u>43.9</u>	43.9	54.9	68.4	105.5	231.5
1.25	<u>24.9</u>	24.9	30.9	38.3	59.9	152.7
1.50	15.0	<u>15.0</u>	18.2	22.3	34.2	90.8
1.75	9.5	<u>9.5</u>	11.3	13.6	20.1	52.0
2.00	<u>6.3</u>	6.3	7.4	8.7	12.4	29.9
2.25	4.4	<u>4.4</u>	5.1	5.8	8.0	17.8
2.50	3.2	<u>3.2</u>	3.7	4.1	5.4	11.1
2.75	2.5	<u>2.5</u>	2.8	3.1	3.9	7.2
3.00	<u>2.0</u>	2.0	2.2	2.4	2.9	5.0
3.25	<u>1.7</u>	1.7	1.8	1.9	2.3	3.6
3.50	<u>1.4</u>	1.4	1.5	1.6	1.9	2.7
3.75	<u>1.3</u>	1.3	1.3	1.4	1.6	2.1
4.00	1.2	<u>1.2</u>	1.2	1.3	1.4	1.7

X VS. X/MR CHARTS

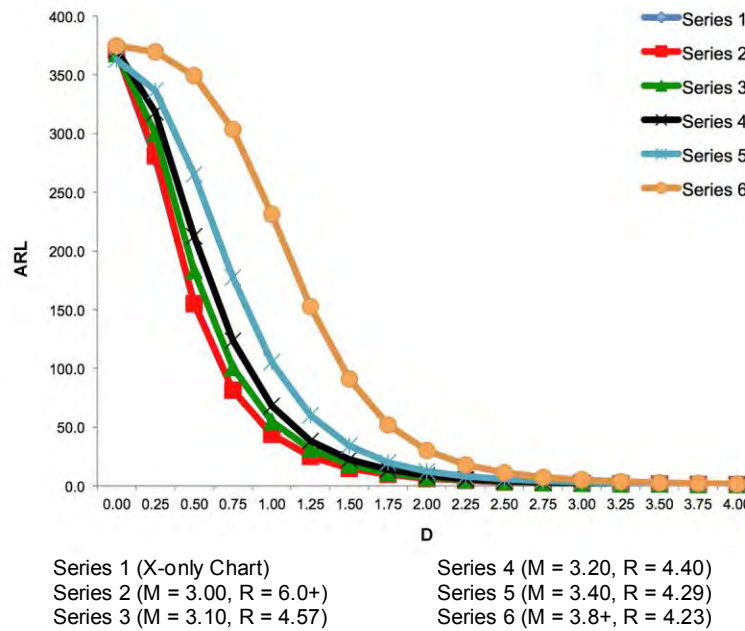


Figure 1. Plotted ARL values of Table 2

Theoretically, the ARLs shown in these 2 “ARL-Columns” should be exactly equal. However, small differences might come from the trapezoidal approximation in Crowder’s published program as well as rounding effects, and also the fact that the normal distribution table used in this work shows only 4 decimals places.

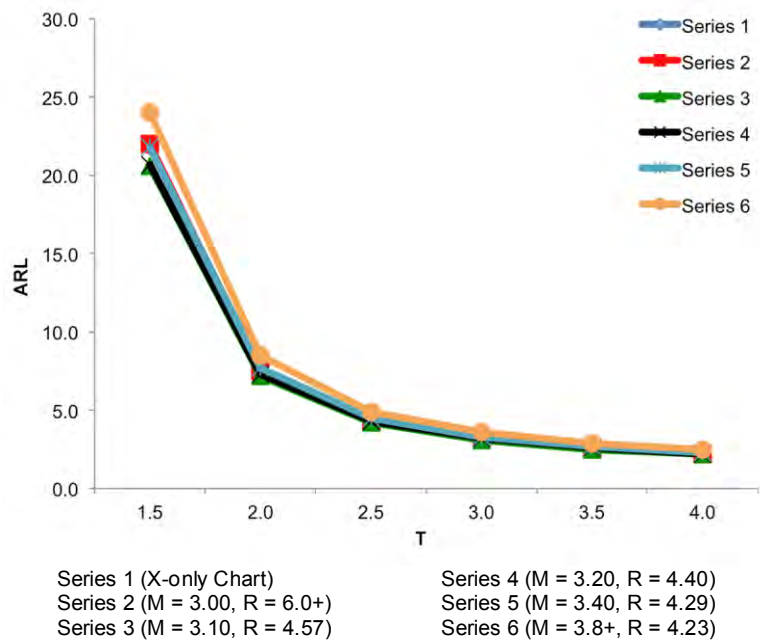
Thus, for Mean Change situation, the X-only Chart is better than the X/MR Chart Combinations since it gives the smallest off-target ARLs (quickest detection of mean process changes).

Sigma Changes Only (Table 3 and Figure 2)

For the second case (Sigma Changes Only), generally, the X/MR Chart with the second smallest M give the smallest off-target ARLs most of the time, except for several large changes in the process standard deviation (i.e., several large T values). For a large increase in process standard deviation (a large T), the Individual Chart alone is better than the X/MR chart Combinations (see Table 3 and Figure 2, for example).

Table 3. Sigma Changes – ARL = 370

T	X-only (L = 3)	X/MR Chart Combinations				
		M = 3.00 R = 6.0+	M = 3.10 R = 4.57	M = 3.20 R = 4.40	M = 3.40 R = 4.29	M = 3.8+ R = 4.23
1.0	370.0	370.0	370.0	370.0	370.0	370.0
1.5	21.9	22.0	<u>20.6</u>	20.7	21.8	24.0
2.0	7.5	7.5	<u>7.2</u>	7.3	7.7	8.5
2.5	4.3	4.3	<u>4.2</u>	4.3	4.5	4.9
3.0	3.2	3.2	<u>3.1</u>	3.2	3.3	3.6
3.5	2.6	2.6	<u>2.5</u>	2.6	2.7	2.9
4.0	2.2	<u>2.2</u>	2.2	2.2	2.3	2.5

**Figure 2.** Plotted ARL values of Table 3. ARL = 370 when $T = 1$ for all series.

Mean and Sigma Change (Table 4 and Figure 3)

For the third case (Both Mean and Sigma Change), an Individual Chart or X/MR Chart Combination with R large (smallest possible M) gives the smallest off-target ARLs. Only a few exceptions appear for $D = 0.5$ with some large T values, where the X/MR Chart Combination with the second smallest M gives the smallest off-target ARL.

X VS. X/MR CHARTS

Where the second smallest M improves on the off-target ARL of the X-only chart, the size of the improvement is clearly quite small. Thus generally speaking, the Individual Chart alone is better than the X/MR Chart Combination (see Table 4 and Figure 3, for example).

Table 4. Mean and Sigma Change – ARL = 370

T	D	X-only (L = 3)	X/MR Chart Combinations				
			M = 3.00 R = 6.0+	M = 3.10 R = 4.57	M = 3.20 R = 4.40	M = 3.40 R = 4.29	M = 3.8+ R = 4.23
1.5	0.0	21.9	22.0	<u>20.6</u>	20.7	21.8	24.0
	0.5	13.7	<u>13.7</u>	14.0	14.9	16.7	20.8
	1.0	6.3	<u>6.2</u>	6.8	7.4	8.9	12.8
	1.5	3.2	<u>3.2</u>	3.5	3.8	4.5	6.5
	2.0	2.0	<u>2.0</u>	2.1	2.3	2.5	3.4
	2.5	<u>1.4</u>	1.4	1.5	1.6	1.7	2.1
	3.0	1.2	<u>1.2</u>	1.2	1.2	1.3	1.5
	3.5	<u>1.1</u>	1.1	1.1	1.1	1.1	1.2
2.0	4.0	1.0	<u>1.0</u>	1.0	1.0	1.0	1.1
	0.0	7.5	7.5	<u>7.2</u>	7.3	7.7	8.5
	0.5	5.5	<u>5.5</u>	5.5	5.7	6.2	7.3
	1.0	3.2	<u>3.2</u>	3.3	3.4	3.8	4.7
	1.5	2.0	<u>2.0</u>	2.1	2.2	2.4	2.8
	2.0	1.4	<u>1.4</u>	1.5	1.5	1.6	1.9
	2.5	1.2	<u>1.2</u>	1.2	1.2	1.3	1.4
	3.0	<u>1.1</u>	1.1	1.1	1.1	1.1	1.2
2.5	3.5	1.0	<u>1.0</u>	1.0	1.0	1.0	1.1
	4.0	1.0	<u>1.0</u>	1.0	1.0	1.0	1.0
	0.0	4.3	4.3	<u>4.2</u>	4.3	4.5	4.9
	0.5	<u>3.5</u>	3.5	3.5	3.6	3.8	4.2
	1.0	2.3	<u>2.3</u>	2.4	2.4	2.6	2.9
	1.5	1.6	<u>1.6</u>	1.6	1.7	1.8	2.0
	2.0	1.3	<u>1.3</u>	1.3	1.3	1.4	1.5
	2.5	1.1	<u>1.1</u>	1.1	1.1	1.1	1.2
3.0	3.0	<u>1.0</u>	1.0	1.0	1.0	1.1	1.1
	3.5	1.0	<u>1.0</u>	1.0	1.0	1.0	1.0
	4.0	1.0	<u>1.0</u>	1.0	1.0	1.0	1.0
	0.0	3.2	3.2	<u>3.1</u>	3.2	3.3	3.6
	0.5	2.7	2.7	<u>2.7</u>	2.7	2.9	3.1
	1.0	2.1	<u>1.9</u>	1.9	2.0	2.1	2.3
	1.5	1.5	<u>1.4</u>	1.5	1.5	1.5	1.7
	2.0	1.2	<u>1.2</u>	1.2	1.2	1.2	1.3
3.0	2.5	1.1	<u>1.1</u>	1.1	1.1	1.1	1.1
	3.0	1.0	<u>1.0</u>	1.0	1.0	1.0	1.0
	3.5	1.0	<u>1.0</u>	1.0	1.0	1.0	1.0
	4.0	<u>1.0</u>	1.0	1.0	1.0	1.0	1.0

T	D	X-only (L = 3)	X/MR Chart Combinations				
			M = 3.00 R = 6.0+	M = 3.10 R = 4.57	M = 3.20 R = 4.40	M = 3.40 R = 4.29	M = 3.8+ R = 4.23
3.5	0.0	2.6	2.6	<u>2.5</u>	2.6	2.7	2.9
	0.5	2.2	<u>2.2</u>	2.2	2.3	2.4	2.6
	1.0	1.9	<u>1.7</u>	1.7	1.7	1.8	2.0
	1.5	1.4	<u>1.3</u>	1.4	1.4	1.4	1.5
	2.0	1.1	<u>1.1</u>	1.2	1.2	1.2	1.2
	2.5	1.1	<u>1.1</u>	1.1	1.1	1.1	1.1
	3.0	1.0	<u>1.0</u>	1.0	1.0	1.0	1.0
	3.5	<u>1.0</u>	1.0	1.0	1.0	1.0	1.0
	4.0	<u>1.0</u>	1.0	1.0	1.0	1.0	1.0
	4.5						
4.0	0.0	2.2	<u>2.2</u>	2.2	2.2	2.3	2.5
	0.5	2.0	<u>2.0</u>	2.0	2.0	2.1	2.2
	1.0	1.8	<u>1.6</u>	1.6	1.6	1.7	1.8
	1.5	1.3	<u>1.3</u>	1.3	1.3	1.3	1.4
	2.0	1.1	<u>1.1</u>	1.1	1.1	1.1	1.2
	2.5	1.0	<u>1.0</u>	1.0	1.0	1.1	1.1
	3.0	1.0	<u>1.0</u>	1.0	1.0	1.0	1.0
	3.5	1.0	<u>1.0</u>	1.0	1.0	1.0	1.0
	4.0	<u>1.0</u>	1.0	1.0	1.0	1.0	1.0
	4.5						

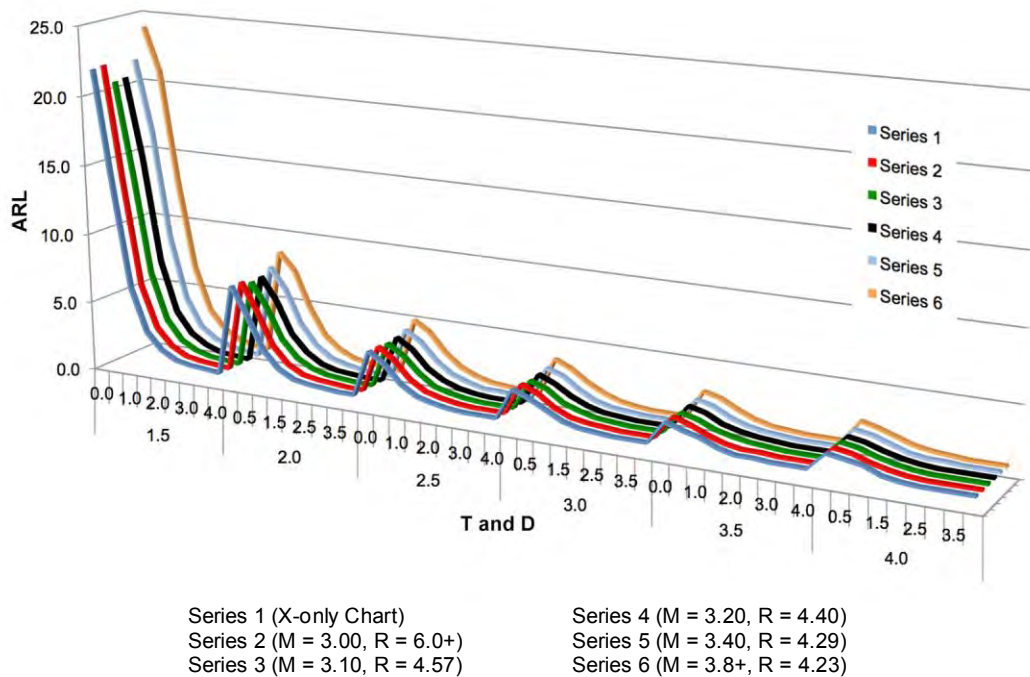


Figure 3. Plotted ARL values of Table 4

Conclusion

It has been found that X chart alone (X-only Chart) is better than the Individuals and Moving Range Chart Combinations (X/MR Chart Combinations) in detecting changes in the process mean. The Individual Chart alone gives smaller off-target ARLs for detecting changes in the process mean.

For the case where the process standard deviation changes, most of the time, X/MR Chart Combinations (with the 2nd smallest M) are better than the Individual Chart alone. Only for small “all OK” ARL values (ARL = 50 and 100), is the Individual Chart alone better than the X/MR Chart Combinations for large T . Specifically, for “all OK” (in-control) ARL = 50, the X-only Chart is better than X/MR Chart Combination when $T > 2.5$. Also for “all OK” (in-control) ARL = 100, the Individual Chart alone gives smaller off-target ARLs than X/MR Chart Combinations when $T > 3.0$. For large “all OK” (in-control) ARLs (ARL = 250, 370, 500, 750, and 1000), X/MR Chart Combination are better than the Individual Chart alone except for ARL = 370 and 750 with $T = 4.0$, where the Individual Chart alone is better.

Finally, the case where both process mean and process standard deviation change, most of the time the X-only Chart is better than the X/MR Chart Combinations. The X/MR Chart Combination can be better than the X-only Chart only when $D = 0.5$ for some values of large T .

Although the X-only chart can be better, the improvement in the off-target ARLs for the last two cases as described above is not really significant (coming in only the 3rd decimal place or beyond). If we round-off the results to the closest integer value, both results will typically be rounded to the same value. Therefore, in general, we can say that X/MR Chart Combination is “nearly” as efficient as the X-chart alone in detecting changes in the process standard deviation. Also, in general, with the same reason as above, the two monitoring schemes are “about” equally effective detecting changes in both process mean and process standard deviation (case 3).

Recommendation

If changes in only the process mean are of concern, it is definitely better to use the X-only Chart monitoring scheme.

If increases in the process standard deviation are the only ones of concern, the recommendation is to use the X-only Chart (for simplicity) even though the X/MR Chart Combinations is “nearly” as efficient as the Individual Chart alone.

Similarly where one is concerned both process mean and standard deviation changes, the recommendation is to use X-only Chart.

Acknowledgement

The author would like to thank the Editor and the referees for their constructive comments and suggestions that improved the presentation of this manuscript. This research represents the authors own work and opinion. It does not reflect any policy nor represent the official position of the U.S. Department of Defense.

References

- Adke, S. R. & Hong, X. (1997). A Supplementary Test Based on the Control Chart for Individuals.” *Journal of Quality Technology*, 29(1), 16-20.
- Amin, R. W. & Ethridge, R. A. (1998). A Note on Individual and Moving Range Control Charts. *Journal of Quality Technology*, 30(1), 70-74.
- Crowder, S. V. (1987a). Computation of ARL for Combined Individual Measurement and Moving Range Charts. *Journal of Quality Technology*, 19(2), 98-102.
- Crowder, S. V. (1987b). A Program for the Computation of ARL for Combined Individual Measurement and Moving Range Charts. *Journal of Quality Technology*, 19(2), 103-106.
- Duncan, A. J. (1974). *Quality Control and Industrial Statistics*. Homewood, IL: Irwin.
- Marks, N. B., & Krehbiel, T. C. (2009). Design and Application of Individuals and Moving Range Control Charts. *The Journal of Applied Business Research*, 25(5), 31-40.
- Radson, D., & Alwan, L. C. (1995). Detecting Variance Reductions Using the Moving Range. *Quality Engineering*, 8(1), 165-178.

X VS. X/MR CHARTS

Rigdon, S. E., Cruthis, E. N., & Champ C. W. (1994). Design Strategies for Individuals and Moving range Control Charts. *Journal of Quality Technology*, 26(4), 274-287.

Roes, K. C. B., Does, R. J. M. M., & Schurink, Y. (1993). Shewhart-Type Control Charts for Individual Observations. *Journal of Quality Techonology*, 25(3), 188-198.

Vardeman, S.B., & Jobe, J. M. (1999). *Statistical Quality Assurance Methods for Engineers*. New York: John Wiley and Sons.

Reliability Estimates of Generalized Poisson Distribution and Generalized Geometric Series Distribution

Adil H. Khan
University of Kashmir
Srinagar, India

T. R. Jan
University of Kashmir
Srinagar, India

Discrete distributions have played an important role in the reliability theory. In order to obtain Bayes estimators, researchers have adopted various conventional techniques. Generalizing the results of Maiti (1995), Chaturvadi and Tomer (2002) dealt with the problem of estimating $P\{X_1, X_2, \dots, X_k \leq Y\}$, where random variables X and Y were assumed to follow a negative binomial distribution. Agit et al. obtained Bayesian estimates of the reliability functions and $P\{X_1, X_2, \dots, X_k \leq Y\}$ considering X and Y following binomial and Poisson distributions. The reliability function of the generalized Poisson and generalized geometric distribution is investigated. The expression for $P\{X_1, X_2, \dots, X_k \leq Y\}$ was obtained with X 's and Y following a Poisson distribution and some particular cases are shown.

Keywords: Generalized Poisson distribution generalized geometric distribution, reliability function, Bayes estimators

Introduction

Much research exists in the literature for estimating various parametric functions of several discrete distributions through classical and Bayesian approaches. Cacoullos and Charalambides (1975) obtained MVUE for truncated binomial and negative binomial distributions. Bayesian estimation of the parameter of binomial distribution has been considered by Chew (1971). Barton (1961) and Glasser (1962) obtained UMVUE of $P(X=x)$ for Poisson distribution. Blyth (1980) studied the absolute error of UMVUE of the probability of success of binomial distribution. For a random variable X following binomial distribution, Pulskanp (1990) has shown that the UMVUE of $P(X=x)$ is admissible under squared-error loss function

Adil H. Khan is a Research Scholar in the Post-Graduate Department of Statistics. Email him khanadil_192@yahoo.com. T. R. Jan is an Assistant Professor in the Post-Graduate Department of Statistics. Email him at drtrjan@gmail.com.

when $X = 0$ or n . Irony (1992) developed Bayesian estimation procedures related to Poisson distribution. Guttman (1958) and Patil (1963) provided UMVUEs of parametric functions of negative binomial distribution. Patil & Wani (1966) obtained UMVEUs of distribution function of various distributions. Roy & Mitra (1957) considered the problem of minimum variance unbiased estimation of univariate power series distribution. Patil (1978) generalized their results to multivariate modified power series distribution. Patil & Bildikar (1966) derived MVUE for logarithmic series distribution.

Discrete distributions have played important role in reliability theory. Kumar & Bhattacharya (1989) considered negative binomial distribution as the life time modal and obtained UMVUEs of the mean life and reliability function. Another measure of reliability is under stress-strength setup in the probability $Pr\{X \leq Y\}$, under the assumption that X and Y followed geometric distribution and derived UMVUE & Bayes estimator. Chaturvedi & Tomer (2002) considered classical & Bayesian Estimation procedures for the reliability function of the negative binomial distribution from a different approach generalizing the results of Maiti (1995), they dealt with the problem of estimating $P\{X_1, X_2, \dots, X_k \leq Y\}$, where random variables X & Y were assumed to follow negative binomial distributions. Chaturvadi, et al. (2007) considered Binomial and Poisson distribution and obtained the Bayesian estimators of reliability function and dealt with the problem of estimating $P\{X_1, X_2, \dots, X_k \leq Y\}$, where the random variables X & Y were assumed to follow binomial and Poisson distributions.

In order to obtain Bayes estimators of parameter and various parametric functions of different distributions, researchers have adopted a conventional technique, i.e., obtaining their posterior means. This article considers the Generalized Poisson and Generalized Geometric distributions and the problems of estimating reliability functions and $P = P\{X_1, X_2, \dots, X_k \leq Y\}$ from a Bayesian viewpoint. Bayes estimators of these parametric functions are derived. It is worth mentioning that in contrary to conventional approach, only estimators of factorial moments are needed to estimate these parametric functions and no separate dealing is needed.

Generalized Poisson Distribution

The random variable follows Generalized Poisson distribution with parameter λ and β if its pmf is

$$P(X = x) = \frac{\lambda^x (1 + x\beta)^{x-1} e^{-\lambda(\beta x + 1)}}{x!}, \quad x = 0, 1, 2, \dots, \lambda > 0, 0 < \beta < \frac{1}{\lambda} \quad (1)$$

The reliability function at a specific mission time, for example, $t_0 (\geq 0)$ is

$$R(t_0) = \sum_{x=t_0}^{\infty} \frac{\lambda^x (1 + x\beta)^{x-1} e^{-\lambda(\beta x + 1)}}{x!} \quad (2)$$

and the hazard rate function is

$$h(t_0) = \frac{P(t_0, \lambda, \beta)}{R(t_0)} = \left[\sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda\beta x} (1 + (x + t_0)\beta)^{x+t_0-1}}{(1 + t_0\beta)^{t_0-1} (x + t_0)_{(x)}} \right]^{-1} \quad (3)$$

Let $\{X_i\}$, $i = 1, 2, 3, \dots, k$ be k independent random variables following a generalized Poisson distribution (1) with parameters λ_i and β (known) and Y is a random variable, independent of X 's following generalized Poisson distribution with parameter u . Denoting

$$X^* = \sum_{i=1}^k X_i \quad \text{and} \quad \lambda^* = \sum_{i=1}^k \lambda_i \quad (4)$$

From additive property of generalized Poisson distribution

$$P = \sum_{x^*=0}^{\infty} \sum_{y=x^*}^{\infty} p(x^*; \lambda^*) \cdot p(y; u) \quad (5)$$

Next, Bayes estimators of $R(t_0)$ and 'P' for generalized Poisson distribution are estimated.

Bayes Estimation Of $R(t_0)$ and 'P' For Generalized Poisson Distribution

The likelihood function given the random sample information $X = (X_1, X_2, \dots, X_n)$ is

$$L(\lambda, \beta / t) = \prod_{i=1}^n \left(\frac{(1 + x_i \beta)^{x_i - 1}}{x_i!} \right) \lambda^t e^{-\lambda(\beta t + n)} \quad (6)$$

where

$$t = \sum_{i=1}^n x_i$$

Because, $\lambda > 0$ consider the prior distribution for λ when β is known to be gamma with parameters (α, θ) and pdf

$$g(\lambda) = \frac{\theta^\alpha e^{-\theta\lambda} \lambda^{\alpha-1}}{\Gamma(\alpha)}, \lambda, \alpha, \theta > 0 \quad (7)$$

From (6) and (7), the posterior density function of λ is given by

$$\Pi(\lambda / t) = \frac{e^{-\lambda(\beta t + n + \theta)} \lambda^{t + \alpha - 1} (\beta t + n + \theta)^{t + \alpha}}{\Gamma(t + \alpha)} \quad (8)$$

The Bayesian estimator of λ^p , for $p > 0$, is given by

$$\begin{aligned} \hat{\lambda}^p &= \int_0^\infty \lambda^p \Pi(\lambda / t) d\lambda \\ \hat{\lambda}^p &= \frac{\Gamma(t + \alpha + p)}{\Gamma(t + \alpha)} (\beta t + n + \theta)^{-p} \end{aligned} \quad (9)$$

Now, Equation (1) can be written as

$$P(x = x) = \frac{1}{x!} \sum_{r=0}^{\infty} (-1)^r \frac{(\beta x + 1)^{r+x-1}}{r!} \lambda^{r+x}$$

On using (9) the Bayes estimator of $P(x;\lambda)$ at a specific point 'X' is

$$\begin{aligned}\hat{p}(x;\lambda) &= \frac{1}{x!} \sum_{r=0}^{\infty} (-1)^r \frac{(\beta x + 1)^{r+x-1}}{r!} \frac{\Gamma(t + \alpha + r + x)}{\Gamma(t + \alpha)} (\beta t + n + \theta)^{-(r+x)} \\ &= \frac{(\beta x + 1)^{x-1}}{(\beta t + n + \theta)^x} \cdot \frac{1}{x!} \sum_{r=0}^{\infty} (-1)^r \frac{(\beta x + 1)^r}{r!} \frac{(t + \alpha + r + x - 1)!}{(t + \alpha - 1)! (\beta t + n + \theta)^r} \quad (10) \\ &= (\beta x + 1)^{x-1} \binom{t + \alpha + x - 1}{x} \frac{(\beta t + n + \theta)^{t+\alpha}}{(\beta(t+x) + n + \theta + 1)^{t+\alpha+x}}\end{aligned}$$

Using (10) in (2), in order to obtain Bayesian estimator of $R(t_0)$, results in

$$\hat{R}(t_0) = \left(\frac{\beta t + n + \theta}{\beta(t+x) + n + \theta + 1} \right)^{t+\alpha} \sum_{x=t_0}^{\infty} \binom{t + \alpha + x - 1}{x} \frac{(\beta x + 1)^{x-1}}{(\beta(t+x) + n + \theta + 1)^x} \quad (11)$$

Also, for obtaining Bayesian estimator for 'P' we consider independent priors for λ^* and u to be gamma with parameters (α_1, θ_1) and (α_2, θ_2) respectively and using equations (4) and (10) is

$$\begin{aligned}\hat{P} &= \frac{\sum_{x^*=0}^{\infty} \sum_{y=x^*}^{\infty} (\beta x^* + 1)^{x^*-1} (\beta y + 1)^{y-1} \binom{t_1 + \alpha_1 + x^* - 1}{x^*} \binom{t_2 + \alpha_2 + y - 1}{y}}{\left(\beta(t_1 + x^*) + \sum_{i=1}^k n + \theta_1 + 1 \right)^{t_1 + \alpha_1 + x^*} (\beta(t_2 + y) + m + \theta_2 + 1)^{t_2 + \alpha_2 + y}} \quad (12)\end{aligned}$$

Special Cases

For $\beta = 0$, equation (1) reduces to Poisson distribution, therefore for $\beta = 0$, equations (11) and (12) give the Bayesian estimators for $R(t_0)$ and P (see Chaturvedi, et al., 2007) and are

$$\hat{R}_1(t_0) = \left(\frac{n + \theta}{n + \theta + 1} \right)^{t + \alpha} \sum_{x=t_0}^{\infty} \binom{t + \alpha + x - 1}{x} \frac{1}{(n + \theta + 1)^x}$$

$$\hat{P} = \frac{\sum_{x^*=0}^{\infty} \sum_{y=x^*}^{\infty} \binom{t_1 + \alpha_1 + x^* - 1}{x^*} \binom{t_2 + \alpha_2 + y - 1}{y} \left(\sum_{i=1}^k n + \theta_i \right)^{t_1 + \alpha_1} (m + \theta_2)^{t_2 + \alpha_2}}{\left(\sum_{i=1}^k n + \theta_i + 1 \right)^{t_1 + \alpha_1 + x^*} (m + \theta_2 + 1)^{t_2 + \alpha_2 + y}}$$

Generalized Geometric Distribution

The random variable 'X' follows Generalized Geometric distribution with parameters α and β if its pmf is

$$P(x=x) = \frac{1}{\beta x + 1} \binom{\beta x + 1}{x} \alpha^x (1 - \alpha)^{1 + (\beta - 1)x} \quad (13)$$

$$0 < \alpha < 1, |\alpha\beta| < 1, x = 0, 1, 2, \dots$$

The reliability function at a specific mission time, for example, $t_0 (\geq 0)$ is

$$R(t_0) = \sum_{x=t_0}^{\infty} \frac{1}{\beta x + 1} \binom{\beta x + 1}{x} \alpha^x (1 - \alpha)^{1 + (\beta - 1)x} \quad (14)$$

and the hazard rate function is

$$h(t_0) = \frac{P(t_0, \alpha, \beta)}{R(t_0)}$$

Bayes Estimation Of $R(t_0)$ For Generalized Geometric Distribution

The likelihood function given the random sample information $X = (X_1, X_2, \dots, X_n)$ is

$$L(\alpha, \beta / t) = \prod_{i=1}^n \left(\frac{1}{\beta x_i + 1} \binom{\beta x_i + 1}{x_i} \right) \alpha^t (1 - \alpha)^{n + (\beta - 1)t} \quad (15)$$

where

$$t = \sum_{i=1}^n x_i$$

Because, $0 < \alpha < 1$, it is assumed that the prior information about α when β is known from Beta distribution with pdf

$$g(\alpha) = \frac{\alpha^{a-1} (1 - \alpha)^{b-1}}{B(a, b)}, 0 < \alpha < 1, a > 0, b > 0 \quad (16)$$

The posterior distribution from (15) and (16) can be written as

$$\Pi(\alpha / t) = \frac{\alpha^{t+a-1} (1 - \alpha)^{n + (\beta - 1)t + b - 1}}{B(t + a, n + (\beta - 1)t + b)} \quad (17)$$

RELIABILITY ESTIMATES OF TWO GENERALIZED DISTRIBUTIONS

The Bayesian estimator of α^p , for $p > 0$, is given by

$$\begin{aligned}\hat{\alpha}^p &= \int_0^{\infty} \alpha^p \prod(\alpha/t) d\alpha \\ \hat{\alpha}^p &= \frac{B(t+a+p, n+(\beta-1)t+b)}{B(t+a, n+(\beta-1)t+b)}\end{aligned}\tag{18}$$

Now, equation (13) can be written as

$$P(x=x) = \frac{1}{\beta x+1} \binom{\beta x+1}{x} \sum_{r=0}^{\infty} (-1)^r \binom{\beta x+1-x}{r} \alpha^{x+r}$$

On using (18) the Bayes estimator of $P(x;\lambda)$ at a specific point 'X' is

$$\begin{aligned}\hat{P}(x;\alpha) &= \frac{1}{\beta x+1} \binom{\beta x+1}{x} \sum_{r=0}^{\infty} (-1)^r \binom{\beta x+1-x}{r} \\ &\quad \times \frac{B(t+a+x+r, n+(\beta-1)t+b)}{B(t+a, n+(\beta-1)t+b)} \\ &= \frac{1}{\beta x+1} \binom{\beta x+1}{x} \frac{B(t+a+x, n+1+(\beta-1)t+b)}{B(t+a, n+1+(\beta-1)t+b)}\end{aligned}\tag{19}$$

Using (19) in (14), in order to obtain Bayesian estimator of $R(t_0)$, results in

$$\hat{R}(t_0) = \sum_{x=t_0}^{\infty} \frac{1}{\beta x+1} \binom{\beta x+1}{x} \frac{B(t+a+x, n+1+(\beta-1)t+b)}{B(t+a, n+1+(\beta-1)t+b)}.$$

References

- Barton, D. E. (1961). Unbiased estimation of a set of probability. *Biometrika*, 48, 227-229.
- Blyth, C. R. (1980). Expected absolute error of the usual estimator of the binomial parameter. *American Statistician*, 34, 155-157.
- Cacoullos, T. & Charalambides, C. (1975). On minimum variance unbiased estimation for truncated binomial and negative binomial distributions. *Annals of the Institute of Statistical Mathematics*, 27, 235-244.
- Chaturvadi, A., Tiwari, N., & Kumar, S. (2007). Some remarks on classical and Bayesian reliability estimation of binomial and Poisson distributions. *Statistical Papers*, 48, 683-693.
- Chaturvedi, A. & Tomer, S. K. (2002). Classical and Bayesian Reliability estimation of the negative binomial distribution. *Journal of Applied Statistical Science*, 11(1), 33-43.
- Chew, V. (1971). Point estimation of the parameter of the binomial distribution. *American Statistician*, 25, 47-50.
- Glasser, G. J. (1962). Minimum variance unbiased estimators for Poisson probabilities. *Technometrics*, 4, 409-418.
- Guttman, I. (1958). A note on series solution of a problem in estimation. *Biometrika*, 45, 565-567.
- Irony, T. Z. (1992). Bayesian estimation for discrete distribution. *Journal of Applied Statistics*, 19, 38-47.
- Kumar, S. & Bhattacharaya, S. K. (1989). Reliability estimation for negative binomial distribution. *Assam Statistical Review*, 3, 104-107.
- Maiti, S. S. (1995). Estimation of $Pr\{X \leq Y\}$ in the Geometric case. *Journal of the Indian Statistical Association*, 33, 87-91.
- Patil, G. P. (1963). Minimum Variance unbiased estimation of generalized Poisson distribution and certain problems of additive number theory. *Annals of Mathematical Statistics*, 34, 1050-1056.
- Patil, G. P. & Bildikar, S. (1966). On minimum variance unbiased estimation for the logarithmic series distribution. *Sankhya*, A28, 239-250.
- Patil, G. P. & Wani, J. K. (1966). Minimum variance unbiased estimation of the discretion function admitting a sufficient statistics. *Annals of Mathematical Statistics*, 18, 39-47.

RELIABILITY ESTIMATES OF TWO GENERALIZED DISTRIBUTIONS

Patil, S. R. (1978). Minimum variance unbiased estimation of multivariate modified power series distribution. *Metrika*, 25, 155-161.

Pulskamp, R. (1990). A note on the estimation of binomial probabilities. *American Statisticians*, 44, 293-295.

Roy, J. & Mitra, S. K. (1957). Unbiased minimum variance estimation in a class of discrete distributions. *Sankhya*, 18, 371-378.

Estimation of Multi Component Systems Reliability in Stress-Strength Models

Adil Hamid Khan
University of Kashmir
Srinagar, India

T.R. Jan
University of Kashmir
Srinagar, India

In a system with standby redundancy, there are a number of components only one of which works at a time and the other remain as standbys. When an impact of stress exceeds the strength of the active component, for the first time, it fails and another from standbys, if there is any, is activated and faces the impact of stresses, not necessarily identical as faced by the preceding component and the system fails when all the components have failed. Sriwastav and Kakaty (1981) assumed that the components stress-strengths are similarly distributed. However, in general the stress distributions will be different from the strength distributions not only in parameter values but also in forms, because stresses are independent of strengths and the two are governed by different physical conditions. Assume the components in the system for both stress and strength are independent and follow different probability distributions viz. Exponential, Gamma, Lindley. Different conditions for stress and strength were considered. Under these assumptions the reliabilities of the system have been obtained with the help of the particular forms of density functions of n-standby system when all stress-strengths are random variables. The expressions for the marginal reliabilities $R(1)$, $R(2)$, $R(3)$ etc. have been obtained based on its stress-strength models. Results obtained by J. Gogoi and M. Bohra are particular case presentations.

Keywords: Reliability, standby redundancy, exponential distribution, Lindley distribution.

Introduction

The reliability of a system is the probability that when operating under stated environmental conditions, the system will perform its intended function adequately. For stress-strength models both the strength of the system, X , and the stress, Y , imposed on it by its operating environments are considered to be random variables.

Adil H. Khan is a Research Scholar in the Post-Graduate Department of Statistics. Email him khanadil_192@yahoo.com. T. R. Jan is an Assistant Professor in the Post-Graduate Department of Statistics. Email him at drtrjan@gmail.com.

The reliability, R , of the system is the probability that the system is strong enough to overcome the stress imposed on it, that is to say $R = \Pr(X > Y)$.

In a standby system, that is, a system with standby redundancy, there are number of components only one of which works at a time and the other remains as standby. When an impact of stress exceeds the strength of the active component, for the first time, it fails and the another component from standbys, if there is any, is activated and faces the impact of stresses, not necessarily identical as faced by the preceding component. The system fails when all the components have failed. This problem has a long history starting with the pioneering work of Birnbaum (1956) and Birnbaum and McCarty (1958). The term stress-strength was first introduced by Church and Harris (1970). Since then significant amount of work has been done both from parametric and non-parametric point of view. A comprehensive treatment of the different stress-strength models till 2001 can be found in the excellent monograph by Kotz et al. (2003). Some of the recent work on the stress-strength model can be obtained in Kundu and Gupta (2005, 2006), Raqab and Kundu (2005), Krishnamoorthy et al. (2007). Gogoi, Borah and Sriwastav (2010) gave an interference model with number of stress a Poisson Process. Gogoi and Borah (2012) obtained the reliability expressions when the stress-strength of the components follow different distributions.

It has been assumed the components stress-strengths are similarly distributed but in general the stress distributions will be different from the strength distributions not only in parameter values but also in forms because stresses are independent of strengths and the two are governed by different physical conditions. It is assumed here that stress-strengths of all the components in the system are independent and obtained the stress-strength parameter $R = \Pr(X > Y)$, when X and Y are independent. Note that the stress-strength parameter plays an important role in the reliability analysis. For example if X is the strength of a system which is subjected to stress Y , then the parameter R measures the system performance and it is very common in the context of mechanical reliability of a system. Moreover, R provides the probability of a system failure, if the system fails whenever the applied stress is greater than its strength.

Model

Consider an n -standby system in which, initially, there are n components, out of which only one is working under impact of stresses and the remaining $(n-1)$ are standby. Whenever the working component fails, one from the standby takes its place and is subjected to impact of stresses and the system works. The system fails when the entire component fails.

Let X_1, X_2, \dots, X_n be a set of n independent random variables, representing the strengths of n components arranged in order of activation in the system and let Y_1, Y_2, \dots, Y_n , be another set of independent random variables representing the stresses on the n components respectively, then the system reliability R_n of the system is given by

$$R_n = R(1) + R(2) + \dots + R(n) \quad (1)$$

where the marginal reliability $R(r)$ is the contribution to the reliability of the system by the r^{th} component and is defined as

$$R(r) = \Pr[X_1 < Y_1, X_2 < Y_2, \dots, X_{r-1} < Y_{r-1}, X_r \geq Y_r]$$

and if $f_i(x)$ and $h_i(y)$ are the probability density functions of X_i and $Y_i, i = 1, 2, 3, \dots, n$ respectively then

$$R(r) = \left[\int_{-\infty}^{\infty} F_1(y) h_1(y) dy \right] \left[\int_{-\infty}^{\infty} F_2(y) h_2(y) dy \right] \dots \left[\int_{-\infty}^{\infty} F_{r-1}(y) h_{r-1}(y) dy \right] \left[\int_{-\infty}^{\infty} \bar{F}_r(y) h_r(y) dy \right] \quad (2)$$

where $F_i(y)$ is the commutative distribution function of X_i and

$$\bar{F}_i(y) = 1 - F(x) \quad (3)$$

Assume that strength and stress follow different distributions, the following cases are considered.

- (I) One parameter exponential strength and three parameter exponential stress.
- (II) Two parameter Lindley strength and one parameter gamma stress.
- (III) Two parameter Lindley strength and two parameter gamma stress.

One parameter exponential strength and three parameter exponential stress

Let $f_i(x)$ be the one parameter exponential strength with parameter λ_i and $h_i(x)$ be the three parameter exponential stress with parameters α_i, β_i and $\mu_i; i = 1, 2, \dots, n$, then

$$f_i(x, \lambda) = \begin{cases} \lambda_i e^{-\lambda_i x_i}; & x_i \geq 0, \lambda_i \geq 0 \\ 0; & \text{otherwise} \end{cases} \quad (4)$$

and

$$h_i(y, \alpha, \beta, \mu) = \begin{cases} \frac{\alpha_i}{\beta_i} \left(1 - e^{-(y_i - \mu_i)/\beta_i}\right); & y_i > \mu_i, \alpha_i > 0, \beta_i > 0 \\ 0; & \text{otherwise} \end{cases} \quad (5)$$

Then from equation (2)

$$R(1) = \int_{\mu_1}^{\infty} \bar{F}_1(y) h_1(y) dy = \frac{\alpha_1 e^{-\lambda_1 \mu_1} \Gamma(\alpha_1) \Gamma(\lambda_1 \beta_1 + 1)}{\Gamma(\alpha_1 + \lambda_1 \beta_1 + 1)}$$

$$\begin{aligned} R(2) &= \left[\int_{\mu_1}^{\infty} F_1(y) h_1(y) dy \right] \left[\int_{\mu_2}^{\infty} \bar{F}_2(y) h_2(y) dy \right] \\ &= \left[1 - \frac{\alpha_1 e^{-\lambda_1 \mu_1} \Gamma(\alpha_1) \Gamma(\lambda_1 \beta_1 + 1)}{\Gamma(\alpha_1 + \lambda_1 \beta_1 + 1)} \right] \left[\frac{\alpha_2 e^{-\lambda_2 \mu_2} \Gamma(\alpha_2) \Gamma(\lambda_2 \beta_2 + 1)}{\Gamma(\alpha_2 + \lambda_2 \beta_2 + 1)} \right] \end{aligned}$$

$$\begin{aligned} R(3) &= \left[\int_{\mu_1}^{\infty} F_1(y) h_1(y) dy \right] \left[\int_{\mu_2}^{\infty} F_2(y) h_2(y) dy \right] \left[\int_{\mu_3}^{\infty} \bar{F}_3(y) h_3(y) dy \right] \\ &= \left[1 - \frac{\alpha_1 e^{-\lambda_1 \mu_1} \Gamma(\alpha_1) \Gamma(\lambda_1 \beta_1 + 1)}{\Gamma(\alpha_1 + \lambda_1 \beta_1 + 1)} \right] \left[1 - \frac{\alpha_2 e^{-\lambda_2 \mu_2} \Gamma(\alpha_2) \Gamma(\lambda_2 \beta_2 + 1)}{\Gamma(\alpha_2 + \lambda_2 \beta_2 + 1)} \right] \\ &\quad \times \left[\frac{\alpha_3 e^{-\lambda_3 \mu_3} \Gamma(\alpha_3) \Gamma(\lambda_3 \beta_3 + 1)}{\Gamma(\alpha_3 + \lambda_3 \beta_3 + 1)} \right] \end{aligned}$$

In general,

$$R(r) = \left[1 - \frac{\alpha_1 e^{-\lambda_1 \mu_1} \Gamma(\alpha_1) \Gamma(\lambda_1 \beta_1 + 1)}{\Gamma(\alpha_1 + \lambda_1 \beta_1 + 1)} \right] \left[1 - \frac{\alpha_2 e^{-\lambda_2 \mu_2} \Gamma(\alpha_2) \Gamma(\lambda_2 \beta_2 + 1)}{\Gamma(\alpha_2 + \lambda_2 \beta_2 + 1)} \right] \cdots \\ \times \left[1 - \frac{\alpha_{r-1} e^{-\lambda_{r-1} \mu_{r-1}} \Gamma(\alpha_{r-1}) \Gamma(\lambda_{r-1} \beta_{r-1} + 1)}{\Gamma(\alpha_{r-1} + \lambda_{r-1} \beta_{r-1} + 1)} \right] \left[\frac{\alpha_r e^{-\lambda_r \mu_r} \Gamma(\alpha_r) \Gamma(\lambda_r \beta_r + 1)}{\Gamma(\alpha_r + \lambda_r \beta_r + 1)} \right] \quad (6)$$

Particular Case

When in (4) $\alpha_i = 1$, the three parameter exponential distribution becomes two parameter exponential distribution. Then the expression $R(r)$ for one parameter exponential strength and two parameter exponential stress is obtained from (6) by putting $\alpha_i = 1; i = 1, 2, \dots, r$ and is given by

$$R(r) = \left[1 - \frac{e^{-\lambda_1 \mu_1}}{\lambda_1 \beta_1 + 1} \right] \left[1 - \frac{e^{-\lambda_2 \mu_2}}{\lambda_2 \beta_2 + 1} \right] \cdots \left[1 - \frac{e^{-\lambda_{r-1} \mu_{r-1}}}{\lambda_{r-1} \beta_{r-1} + 1} \right] \left[\frac{e^{-\lambda_r \mu_r}}{\lambda_r \beta_r + 1} \right] \quad (7)$$

The result in (7) is obtained by Gogoi and Borah (2012)

Two parameter Lindley strength and one parameter gamma stress

Let $f_i(x)$ be the two parameter Lindley strength with parameters θ_i and α_i and $h_i(x)$ be the one parameter gamma stress with parameter $m_i, i = 1, 2, \dots, n$, the

$$f_i(x, \theta, \alpha) = \begin{cases} \frac{\theta_i^2}{\theta_i + \alpha_i} (1 + \theta_i x_i) e^{-\theta_i x_i}; x_i \geq 0, \theta_i > 0, \alpha_i > -\theta_i \\ 0; \text{ otherwise} \end{cases} \quad (8)$$

and

$$h_i(y, m) = \begin{cases} \frac{1}{\Gamma(m_i)} e^{-y_i} y_i^{m_i}; y_i \geq 0, m_i \geq 1 \\ 0; \text{ otherwise} \end{cases} \quad (9)$$

Then from equation (2)

$$\begin{aligned}
 R(1) &= \int_0^{\infty} \bar{F}_1(y) h_1(y) dy = \frac{(\theta_1 + \alpha_1)(\theta_1 + 1) + \alpha_1 \theta_1 m_1}{(\theta_1 + \alpha_1)(\theta_1 + 1)^{m_1 + 1}} \\
 R(2) &= \left[\int_0^{\infty} F_1(y) h_1(y) dy \right] \left[\int_0^{\infty} \bar{F}_2(y) h_2(y) dy \right] \\
 &= \left[1 - \frac{(\theta_1 + \alpha_1)(\theta_1 + 1) + \alpha_1 \theta_1 m_1}{(\theta_1 + \alpha_1)(\theta_1 + 1)^{m_1 + 1}} \right] \left[\frac{(\theta_2 + \alpha_2)(\theta_2 + 1) + \alpha_2 \theta_2 m_2}{(\theta_2 + \alpha_2)(\theta_2 + 1)^{m_2 + 1}} \right] \\
 R(3) &= \left[\int_0^{\infty} F_1(y) h_1(y) dy \right] \left[\int_0^{\infty} F_2(y) h_2(y) dy \right] \left[\int_0^{\infty} \bar{F}_3(y) h_3(y) dy \right] \\
 &= \left[1 - \frac{(\theta_1 + \alpha_1)(\theta_1 + 1) + \alpha_1 \theta_1 m_1}{(\theta_1 + \alpha_1)(\theta_1 + 1)^{m_1 + 1}} \right] \left[1 - \frac{(\theta_2 + \alpha_2)(\theta_2 + 1) + \alpha_2 \theta_2 m_2}{(\theta_2 + \alpha_2)(\theta_2 + 1)^{m_2 + 1}} \right] \\
 &\quad \times \left[\frac{(\theta_3 + \alpha_3)(\theta_3 + 1) + \alpha_3 \theta_3 m_3}{(\theta_3 + \alpha_3)(\theta_3 + 1)^{m_3 + 1}} \right]
 \end{aligned}$$

In general

$$\begin{aligned}
 R(r) &= \left[1 - \frac{(\theta_1 + \alpha_1)(\theta_1 + 1) + \alpha_1 \theta_1 m_1}{(\theta_1 + \alpha_1)(\theta_1 + 1)^{m_1 + 1}} \right] \left[1 - \frac{(\theta_2 + \alpha_2)(\theta_2 + 1) + \alpha_2 \theta_2 m_2}{(\theta_2 + \alpha_2)(\theta_2 + 1)^{m_2 + 1}} \right] \dots \\
 &\quad \left[1 - \frac{(\theta_{r-1} + \alpha_{r-1})(\theta_{r-1} + 1) + \alpha_{r-1} \theta_{r-1} m_{r-1}}{(\theta_{r-1} + \alpha_{r-1})(\theta_{r-1} + 1)^{m_{r-1} + 1}} \right] \left[\frac{(\theta_r + \alpha_r)(\theta_r + 1) + \alpha_r \theta_r m_r}{(\theta_r + \alpha_r)(\theta_r + 1)^{m_r + 1}} \right]
 \end{aligned} \tag{10}$$

Particular case

1) When $\alpha_i = 1$ in (8) two parameter Lindley distribution becomes one parameter Lindley distribution. Then the expression $R(r)$ for one parameter Lindley strength and one parameter gamma stress is obtained from (10) by putting $\alpha_i = 1; i = 1, 2, \dots, r$ and is given by

$$R(r) = \left[1 - \frac{(\theta_1 + 1)(\theta_1 + 1) + \theta_1 m_1}{(\theta_1 + 1)(\theta_1 + 1)^{m_1 + 1}} \right] \left[1 - \frac{(\theta_2 + 1)(\theta_2 + 1) + \theta_2 m_2}{(\theta_2 + 1)(\theta_2 + 1)^{m_2 + 1}} \right] \cdots \left[1 - \frac{(\theta_{r-1} + 1)(\theta_{r-1} + 1) + \theta_{r-1} m_{r-1}}{(\theta_{r-1} + 1)(\theta_{r-1} + 1)^{m_{r-1} + 1}} \right] \left[\frac{(\theta_r + 1)(\theta_r + 1) + \theta_r m_r}{(\theta_r + 1)(\theta_r + 1)^{m_r + 1}} \right] \quad (11)$$

(The result in (11) is obtained by Gogoi and Borah (2012))

2) When $\alpha_i = 0$ in (8) two parameter Lindley distribution becomes one parameter exponential distribution. Then the expression $R(r)$ for one parameter exponential strength and one parameter gamma stress is obtained from (10) by putting $\alpha_i = 1; i = 1, 2, \dots, r$ and is given by

$$R(r) = \left[1 - \frac{1}{(\theta_1 + 1)^{m_1}} \right] \left[1 - \frac{1}{(\theta_2 + 1)^{m_2}} \right] \cdots \left[1 - \frac{1}{(\theta_{r-1} + 1)^{m_{r-1}}} \right] \left[\frac{1}{(\theta_r + 1)^{m_r}} \right] \quad (12)$$

Two parameter Lindley strength and two parameter gamma stress

Let $f_i(x)$ be the two parameter Lindley strength with parameters θ_i and α_i and $h_i(x)$ be the two parameter gamma stress with parameter m_i and $\mu_i, i = 1, 2, \dots, n$, then

$$f_i(x, \theta, \alpha) = \begin{cases} \frac{\theta_i^2}{\theta_i + \alpha_i} (1 + \theta_i x_i) e^{-\theta_i x_i}; x_i \geq 0, \theta_i > 0, \alpha_i > -\theta_i \\ 0; & \text{otherwise} \end{cases} \quad (13)$$

and

$$h_i(y, m, \mu) = \begin{cases} \frac{1}{\mu_i^{m_i} \Gamma(m_i)} e^{\frac{-y_i}{\mu_i}} y_i^{m_i - 1}; y_i, m_i, \mu_i > 0 \\ 0; & \text{otherwise} \end{cases} \quad (14)$$

Then from equation (2)

$$R(1) = \int_0^{\infty} \bar{F}_1(y) h_1(y) dy = \frac{(\theta_1 + \alpha_1)(\theta_1 \mu_1 + 1) + \alpha_1 \theta_1 m_1 \mu_1}{(\theta_1 + \alpha_1)(\theta_1 \mu_1 + 1)^{m_1 + 1}}$$

$$\begin{aligned} R(2) &= \left[\int_0^{\infty} F_1(y) h_1(y) dy \right] \left[\int_0^{\infty} \bar{F}_2(y) h_2(y) dy \right] \\ &= \left[1 - \frac{(\theta_1 + \alpha_1)(\theta_1 \mu_1 + 1) + \alpha_1 \theta_1 m_1 \mu_1}{(\theta_1 + \alpha_1)(\theta_1 \mu_1 + 1)^{m_1 + 1}} \right] \left[\frac{(\theta_2 + \alpha_2)(\theta_2 \mu_2 + 1) + \alpha_2 \theta_2 m_2 \mu_2}{(\theta_2 + \alpha_2)(\theta_2 \mu_2 + 1)^{m_2 + 1}} \right] \end{aligned}$$

$$\begin{aligned} R(3) &= \left[\int_0^{\infty} F_1(y) h_1(y) dy \right] \left[\int_0^{\infty} F_2(y) h_2(y) dy \right] \left[\int_0^{\infty} \bar{F}_3(y) h_3(y) dy \right] \\ &= \left[1 - \frac{(\theta_1 + \alpha_1)(\theta_1 \mu_1 + 1) + \alpha_1 \theta_1 m_1 \mu_1}{(\theta_1 + \alpha_1)(\theta_1 \mu_1 + 1)^{m_1 + 1}} \right] \left[1 - \frac{(\theta_2 + \alpha_2)(\theta_2 \mu_2 + 1) + \alpha_2 \theta_2 m_2 \mu_2}{(\theta_2 + \alpha_2)(\theta_2 \mu_2 + 1)^{m_2 + 1}} \right] \\ &\quad \times \left[\frac{(\theta_3 + \alpha_3)(\theta_3 \mu_3 + 1) + \alpha_3 \theta_3 m_3 \mu_3}{(\theta_3 + \alpha_3)(\theta_3 \mu_3 + 1)^{m_3 + 1}} \right] \end{aligned}$$

In general,

$$\begin{aligned} R(r) &= \left[1 - \frac{(\theta_1 + \alpha_1)(\theta_1 \mu_1 + 1) + \alpha_1 \theta_1 m_1 \mu_1}{(\theta_1 + \alpha_1)(\theta_1 \mu_1 + 1)^{m_1 + 1}} \right] \left[1 - \frac{(\theta_2 + \alpha_2)(\theta_2 \mu_2 + 1) + \alpha_2 \theta_2 m_2 \mu_2}{(\theta_2 + \alpha_2)(\theta_2 \mu_2 + 1)^{m_2 + 1}} \right] \dots \\ &\quad \left[1 - \frac{(\theta_{r-1} + \alpha_{r-1})(\theta_{r-1} \mu_{r-1} + 1) + \alpha_{r-1} \theta_{r-1} m_{r-1} \mu_{r-1}}{(\theta_{r-1} + \alpha_{r-1})(\theta_{r-1} \mu_{r-1} + 1)^{m_{r-1} + 1}} \right] \left[\frac{(\theta_r + \alpha_r)(\theta_r \mu_r + 1) + \alpha_r \theta_r m_r \mu_r}{(\theta_r + \alpha_r)(\theta_r \mu_r + 1)^{m_r + 1}} \right] \end{aligned} \quad (15)$$

Particular Case

1) When $\alpha_i = 1$ in (13) two parameter Lindley distribution becomes one parameter Lindley distribution. Then the expression $R(r)$ for one parameter Lindley strength and two parameter gamma stress is obtained from (15) by putting $\alpha_i = 1; i = 1, 2, \dots, r$ and is given by

$$R(r) = \left[1 - \frac{(\theta_1 + 1)(\theta_1 \mu_1 + 1) + \theta_1 m_1 \mu_1}{(\theta_1 + 1)(\theta_1 \mu_1 + 1)^{m_1 + 1}} \right] \left[1 - \frac{(\theta_2 + 1)(\theta_2 \mu_2 + 1) + \theta_2 m_2 \mu_2}{(\theta_2 + 1)(\theta_2 \mu_2 + 1)^{m_2 + 1}} \right] \cdots \quad (16)$$

$$\left[1 - \frac{(\theta_{r-1} + 1)(\theta_{r-1} \mu_{r-1} + 1) + \theta_{r-1} m_{r-1} \mu_{r-1}}{(\theta_{r-1} + 1)(\theta_{r-1} \mu_{r-1} + 1)^{m_{r-1} + 1}} \right] \left[\frac{(\theta_r + 1)(\theta_r \mu_r + 1) + \theta_r m_r \mu_r}{(\theta_r + 1)(\theta_r \mu_r + 1)^{m_r + 1}} \right]$$

2) When $\alpha_i = 0$ in (13) two parameter Lindley distribution becomes one parameter exponential distribution. Then the expression $R(r)$ for one parameter exponential strength and two parameter gamma stress is obtained from (15) by putting $\alpha_i = 1; i = 1, 2, \dots, r$ and is given by

$$R(r) = \left[1 - \frac{1}{(\theta_1 \mu_1 + 1)^{m_1}} \right] \left[1 - \frac{1}{(\theta_2 \mu_2 + 1)^{m_2}} \right] \cdots \left[1 - \frac{1}{(\theta_{r-1} \mu_{r-1} + 1)^{m_{r-1}}} \right] \left[\frac{1}{(\theta_r \mu_r + 1)^{m_r}} \right] \quad (17)$$

The results in (16) and (17) were obtained by Gogoi and Borah (2012)

References

- Birnbaum, Z. W. (1956). On a use of the Mann-Whitney Statistics, *Proceedings of the 3rd Berkley symposium I*, 13-17.
- Birnbaum, Z. W., & McCarty, R. C. (1958). A distribution-free upper confidence bound for $\Pr\{Y < X\}$, based on independent samples of X and Y. *The Annals of Mathematical Statistics*, 29(2), 558-562.
- Church, J. D., & Harris, B. (1970). The estimation of reliability from stress-strength relationships. *Technometrics*, 12, 49-54.
- Ghitany, M. E., Atieh, B., & Nadarajah, S. (2008). Lindley distribution and its applications. *Mathematical and Computers in Simulations*, 78, 493-504.
- Gogoi, J., Borah, M., & Sriwastav, G. L. (2010). An interference model with number of stresses a Poisson process. *IAPQR Transactions*, 34(2), 139-152.
- Gogoi, J., & Borah, M. (2012). Estimation of reliability for multi-component systems using exponential Gamma and Lindley stress-strength distributions. *Journal of Reliability and Statistics Studies*, 5(1), 33-41.
- Kakari, M. C. (1983). *Interference theory and reliability*. (Unpublished doctoral dissertation). Dibrugarh University, India.

ESTIMATION OF MULTI COMPONENT SYSTEMS RELIABILITY

Kotz, S., Lumelskii, Y., & Pensky, M. (2003). *The Stress-Strength Model and its Generalizations: Theory and Applications*. Singapore: World Scientific Press.

Krishnamoorthy, K., Shubhabarata, M., & Guo, H. (2007). Inference on Reliability in two parameter exponential stress-strength model, *Metrika*, 65(3), 261-273.

Kundu, D., & Gupta, R. D. (2005). Estimation of $P[Y < X]$ for generalized exponential distribution. *Metrika*, 61, 291–308.

Kundu, D., & Gupta, R. D. (2006). Estimation of $R = P[Y < X]$ for Weibull distributions. *IEEE Transactions on Reliability*, 55, 270–280.

Pandit, S. N. N., & Sriwastav, G. L. (1975). Studies in Cascade Reliability-I, *IEEE Transactions on Reliability*, 24(1), 53-56.

Raqab, M. Z., & Kundu, D. (2005). Comparison of Different Estimators of $P[Y < X]$ for a Scaled Burr Type X Distribution. *Communications in Statistics-Simulation and Computation*, 34, 465–483.

Sriwastav, G. L., & Kakati, M. C. (1981). A stress-strength model with redundancy. *IAPQR Transactions*, 6(1), 21-27.

Double Bootstrap Confidence Interval Estimates with Censored and Truncated Data

Jayanthi Arasan
University Putra Malaysia
Serdang, Selangor, Malaysia

Mohd Bakri Adam
University Putra Malaysia
Serdang, Selangor, Malaysia

Traditional inferential procedures often fail with censored and truncated data, especially when sample sizes are small. In this paper we evaluate the performances of the double and single bootstrap interval estimates by comparing the double percentile (DB-p), double percentile-t (DB-t), single percentile (B-p), and percentile-t (B-t) bootstrap interval estimation methods via a coverage probability study when the data is censored using the log logistic model. We then apply the double bootstrap intervals to real right censored lifetime data on 32 women with breast cancer and failure data on 98 brake pads where all the observations were left truncated.

Keywords: Double bootstrap, censored, simulation, truncated, survival

Introduction

Modeling and data analysis is never complete without reliable statistical inferential procedures such as the confidence intervals or hypothesis testing. These are powerful tools that help us make certain conclusions regarding the population and its parameters based on sample data. The confidence interval can also be used to indicate the reliability of our estimates. However, it is not easy to obtain the exact solutions for some of these inferential procedures especially in cases involving more complex data structures such as incomplete, censored or truncated data. Thus, many have resorted to the much simplified techniques based on the asymptotic normality of the maximum likelihood estimates. In recent years, the bootstrapping techniques have taken over some of these existing methods because they relieve us from certain assumptions based on the asymptotic statistical theory. Thus, the

*Dr. Arasan is a lecturer in the Computational Statistics and Operations Research Laboratory, Institute for Mathematical Research. Email her at jayanthi@upm.edu.my.
Dr. Adam is a lecturer in the Computational Statistics and Operations Research Laboratory, Institute for Mathematical Research. Email him at bakri@upm.edu.my.*

DOUBLE BOOTSTRAP CONFIDENCE INTERVAL ESTIMATES

bootstrap is widely preferred when sample sizes are low or data is censored or truncated since the Wald is known to be highly asymmetrical in these cases with actual coverage probability significantly different from the nominal specification as discussed by Jeng and Meeker (2000), Doganaksoy and Schmee (1993a; 1993b) and Arasan and Lunn (2009).

The bootstrap method as described by Efron and Tibshirani (1993) is a direct application of the plug-in principle which is a way of understanding the population based on estimates from random samples drawn from the population. The standard bootstrap confidence interval is constructed from information obtained directly from the data without any theoretical assumptions. In recent years the double bootstrap has gained a lot of popularity because it typically has a higher order of accuracy. According to McCullough and Vinod (1998), and Letson and McCullough (1998), the double bootstrap enhances the accuracy of the ordinary bootstrap by estimating an error and then using this error to adjust the ordinary bootstrap in order to reduce its error.

Efron (1993) introduced the bootstrap percentile (B-p), percentile-t (B-t), and the bias-corrected percentile (BCa) intervals, see also Efron (1981b; 1981a). Efron (1985) claims that the bootstrap confidence interval reduces most of the error in standard approximation. He also describes some of these intervals as invariant under transformation, thus producing accurate results without involving knowledge of the normalizing transformations. An extensive survey of different bootstrap methods for producing good confidence interval estimates is given in DiCiccio and Efron (1996). Singh (1981) established the second order accuracy of the bootstrap confidence interval by applying Edgeworth theory to the B-t interval. Hall (1986; 1988a; 1988b) examined several different bootstrap interval estimation methods that can be used in both parametric and nonparametric settings and concluded that B-t and BCa methods were superior to other methods. More applied works on bootstrap confidence intervals were done by Arasan and Lunn (2008), Robinson (1983), Schenker (1985), and Jeng and Meeker (2000).

The Model

The log-logistic distribution is very popular in survival studies because it has a hazard rate that increases in the beginning and slowly starts to decrease after a finite time. These types of non-monotonic hazard rates are very popular in medical studies especially those involving lung cancer, breast cancer and kidney or heart transplant patients. This distribution has been studied by various authors such as Bennet (1983) who explored and provided the linear model for the log odds on

survival. Gupta et al. (1999) proved analytically that unique maximum likelihood estimates exist for the parameters of this model and analyzed a lung cancer data. Mazucheli et al. (2005) compared the accuracy of Wald confidence interval with the B-p and B-t intervals for the mode of the hazard function of the log logistic distribution. Other authors who have done significant work using this model are [Cox and Lewis (1966)], [Cox, Oakes, O'Quigley and Struthers (1982)]. The model can also easily be extended to accommodate covariates, truncated data and all types of censored observations such as left, right and interval. More discussions on truncated data can be found in Lawless (1982).

Lifetime data are sometimes truncated due to some of the conditions in the study design. When the lifetime t_i for the i^{th} subject is forced to lie between the interval $[u_i, v_i]$, where u_i and v_i are left and right truncation times respectively, then t_i is said to be either left or right truncated. Subjects who do not experience the event within this window will not be included in the study. So, subjects are left truncated only if they were already at risk before entering the study. So the current lifetime of subject i at selection is $t_i \geq u_i$ where $u_i > 0$ and. Similarly right truncated data are data where $t_i \leq v_i$. Thus, right truncation occurs when all the subjects have already experienced the event of interest when they enter the study. On the other hand right censoring occurs when a subject's event time is unknown due to reasons such as study has ended or subject has left the study. Left censoring occurs when the event of interest has already occurred before the study started. In both cases subjects are still included in the study but it is acknowledged that their event time is above or below a certain point.

There are several different parameterizations for the log logistic distribution. If $-\infty < \delta < \infty$ is the scale parameter and $\beta > 0$ is the shape parameter the density (pdf) and survivor function of the log logistic are

$$f(t, \delta, \beta) = \frac{e^{\delta} t^{\beta-1}}{[1 + e^{\delta} t^{\beta}]^2}$$

$$S(t, \delta, \beta) = \frac{1}{1 + e^{\delta} t^{\beta}}.$$

Suppose we have both right censored and uncensored lifetimes for $i = 1, 2, \dots, n$ observations. Let c_i be the indicator variable assuming the value of 1 if data is uncensored or 0 otherwise. The log-likelihood function for the full sample is given by

DOUBLE BOOTSTRAP CONFIDENCE INTERVAL ESTIMATES

$$\begin{aligned}
 l(\delta, \beta) &= \ln \left\{ \prod_{i=1}^n [f(t_i; \delta; \beta)]^{c_i} [S(t_i; \delta; \beta)]^{(1-c_i)} \right\} \\
 &= \sum_{i=1}^n c_i \left((\beta - 1) \ln(t_i) + \ln(\beta) + \delta - 2 \ln(1 + t_i^\beta e^\delta) \right) \\
 &\quad - (1 - c_i) \ln(1 + t_i^\beta e^\delta)
 \end{aligned} \tag{1}$$

Suppose there is left truncated data for $i = 1, 2, \dots, n$ observations. Let u_i be the left truncation time for the i^{th} subject. The log-likelihood function for the full sample is given by

$$\begin{aligned}
 l(\delta, \beta) &= \ln \left\{ \prod_{i=1}^n \frac{f(t_i; \delta; \beta)}{S(u_i; \delta; \beta)} \right\} \\
 &= \sum_{i=1}^n (\beta - 1) \ln(t_i) + \ln(\beta) + \delta - 2 \ln(1 + t_i^\beta e^\delta) + \ln(1 + u_i^\beta e^\delta)
 \end{aligned} \tag{2}$$

Confidence interval estimates

A bootstrap procedure depends on how the bootstrap sampling is done, namely parametric or nonparametric sampling procedures. In the parametric bootstrap sampling procedure, B bootstrap samples each of size n are generated from the assumed parametric distribution. Then, the bootstrap estimates, $\hat{\theta}_b^*$, $b = 1, 2, \dots, B$ are estimates calculated from each of these bootstrap samples of size n . The nonparametric procedure requires the sampling of a large number of B bootstrap samples with replacement from the original data set with each observation having equal probability of being chosen. This technique of resampling clearly requires the assumption that the data are independent. Following that, the bootstrap estimates are calculated in the same way as described before. In this research we employ the nonparametric bootstrap sampling procedure since we wish to incorporate censored and truncated observations in our dataset.

Single bootstrap without pivot (Percentile interval or B-p)

A clear and thorough understanding of the single bootstrap interval estimation procedures is essential before moving on to any of the double bootstrap methods. If θ is our parameter of interest and $\hat{\theta}$ its estimate using sample data, then $\hat{\theta}_b^*$ is the estimate of θ using the b^{th} bootstrap sample. The B-p method is rather simple and

constructs confidence intervals directly from the percentiles of the bootstrap distribution of the estimated parameters. In this procedure B bootstrap samples, y^b , $b = 1, 2, \dots, B$ will be generated using the nonparametric bootstrap sampling method. Following that, estimate the bootstrap version of the parameter estimates, $\hat{\theta}_b^*$ for each of the bootstrap sample, y^b . The $100(1-\alpha)\%$ percentile interval for θ is

$\{\hat{\theta}_{[l]}^*, \hat{\theta}_{[u]}^*\}$, where, $l = B \times \frac{\alpha}{2}$, $u = B \times \left(1 - \frac{\alpha}{2}\right)$ and $\hat{\theta}_{[\kappa]}^*$, $\kappa = 1, 2, \dots, B$ is the ordered list of the B values of $\hat{\theta}^*$. For ease of computation and accuracy large values of B that give integer values of l and u should preferably be chosen.

The B-p method is said to be transformation-respecting and has the ability to automatically produce accurate results without any normalizing transformations as described in Efron and Tibshirani (1993). Thus it becomes especially useful when the distribution of $\hat{\theta}$ is not approximately normally distributed, since in this case the Wald interval would not perform well unless an appropriate transformation is used.

Single bootstrap with pivot (Percentile-t interval or B-t)

The B-t method involves a bit more work than the B-p interval since it requires the standard error of an estimate. In this method, a bootstrap table consisting the percentiles of the bootstrap version of the standardized values of the parameter estimates (approximate pivot) is constructed using the available data. The property of the approximate pivot whose distribution is approximately the same for all parameter allows the formation of this bootstrap distribution. Following that, this bootstrap table is used to construct the B-t confidence intervals. The main highlight of this method is that it is only dependent on the data in hand and does not require any normal theory assumptions. However, depending on the data available, the bootstrap distribution produced (B-t percentiles) can be asymmetric about 0, which may produce more asymmetrical intervals although at a much better coverage probability.

In this procedure, compute $\hat{\theta}_b^*$ for $b = 1, 2, \dots, B$ bootstrap samples and obtain

$$R_b^* = \frac{\hat{\theta}_b^* - \hat{\theta}}{se(\hat{\theta}_b^*)}. \quad (3)$$

DOUBLE BOOTSTRAP CONFIDENCE INTERVAL ESTIMATES

Here, $se(\hat{\theta}_b^*)$ is the estimated standard error of $\hat{\theta}_b^*$ for the bootstrap sample, y^b . In B-t confidence interval the $\frac{\alpha}{2}$ quantile are based on B-t percentiles which can be obtained using the data. In order to do this obtain $R_{[\kappa]}^*$, $\kappa = 1, 2, \dots, B$ which is the ordered list of the B values of R^* . The $\frac{\alpha}{2}$ percentile of R^* is then the value $R_{[B-\frac{\alpha}{2}]}^*$. Then, the $100(1-\alpha)\%$ confidence interval for θ is $\{\hat{\theta} - R_{[l]}^* se(\hat{\theta}), \hat{\theta} - R_{[u]}^* se(\hat{\theta})\}$ where, $l = B \times (1 - \frac{\alpha}{2})$ and $u = B \times \frac{\alpha}{2}$.

Double bootstrap procedures

The double bootstrap procedure requires resampling from bootstrap samples in order to further reduce the bias and correct the errors in the bootstrap procedures, see Martin (1992). Similarly a double bootstrap confidence interval procedure is a further iteration to the ordinary bootstrap confidence interval procedure which would further reduce the order of magnitude of coverage error. Both B-p and B-t, has under mild regularity conditions, a 2 sided coverage error equals $O(n^{-1})$, at nominal level α . It follows that a further iteration of the ordinary bootstrap confidence interval would further reduce the order of magnitude of coverage error to $O(n^{-2})$.

Double bootstrap without pivot (Double B-p)

The double bootstrap without a pivot or double B-p is given by Shi (1992) and also discussed by Letson and McCullough (1998) and is constructed as follows. First, draw B single bootstrap samples, denoted $y^1, y^2, \dots, y^b, b = 1, 2, \dots, B$. Then, for each b draw another $c = 1, 2, \dots, C$ bootstrap resamples. Following that calculate $\hat{\theta}_{bc}^{**}$ for each double bootstrap samples. In the next step we have to calculate the number of $\hat{\theta}_{bc}^{**}$ that is lesser or equal to $\hat{\theta}$ for each c and divide this number by C

$$Q_b = \frac{\#(\hat{\theta}_{bc}^{**} \leq \hat{\theta})}{C} \quad (4)$$

Following that if the ordered values of Q_b are $Q_{[1]}, Q_{[2]}, \dots, Q_{[B]}$ then the $(1-\alpha)\%$ double percentile bootstrap confidence interval for θ is $\{\hat{\theta}_{[l]}^*, \hat{\theta}_{[u]}^*\}$ where

$l = B.Q_{\left[\frac{\alpha}{2} \times B\right]}$, $u = B.Q_{\left[\left(1-\frac{\alpha}{2}\right) \times B\right]}$ and $\hat{\theta}_{[\kappa]}^*$, $\kappa = 1, 2, \dots, B$ is the ordered list of the B values of $\hat{\theta}^*$. In cases where l and u are not integers, they should be rounded to the nearest integer lesser than or equal to their values.

Double bootstrap with pivot (Double B-t)

The double bootstrap with pivot or double B-t was discussed Mccullough and Vinod (1998) and also Letson and Mccullough (1998). In order to construct the double B-t confidence interval for the parameter, θ , recall that for the B-t interval we need to compute R_b^* as given by (3). Following that we now have to resample C double bootstrap samples from each of the single bootstrap samples and obtain

$$R_{bc}^{**} = \frac{\hat{\theta}_{bc}^{**} - \hat{\theta}_b^*}{se(\hat{\theta}_{bc}^{**})}.$$

In the next step we have to calculate the number of times the second stage root R_{bc}^{**} is lesser or equal to the first stage root R_b^* for each c and divide this number by C

$$Z_b = \frac{\#(R_{bc}^{**} \leq R_b^*)}{C}. \quad (5)$$

Then, if the ordered values of Z_b are $Z_{[1]}, Z_{[2]}, \dots, Z_{[B]}$, the $(1-\alpha)\%$ double B-t confidence interval for θ is $\{\hat{\theta} - R_{[l]}^*.se(\hat{\theta}), \hat{\theta} - R_{[u]}^*.se(\hat{\theta})\}$ where $l = B.Z_{\left[\left(1-\frac{\alpha}{2}\right) \times B\right]}$,

$u = B.Z_{\left[\frac{\alpha}{2} \times B\right]}$ and $R_{[\kappa]}^*$, $\kappa = 1, 2, \dots, B$ is the ordered list of the B values of R^* .

Simulation study

A simulation study was conducted using $N = 1000$ samples of size $n = 25, 30, 40$ and 50 to compare the performance of the confidence interval estimates discussed in the previous section for the parameters of the log logistic model with censored data. We used $\alpha = 0.05$ and $\alpha = 0.10$ where α is the nominal error probability and censoring proportion (cp) of 10% and 15% . We compared a total of 4 methods

namely the B-p, B-t, DB-p and DB-t. Following that, we calculated the estimated coverage probability error by adding the number of times in which an interval did not contain the true parameter value divided by the total number of samples.

The estimated left (right) error probability was calculated by adding the number of times the left (right) endpoint was more (less) than the true parameter value divided by the total number of samples. Following Doganaksoy and Schmee (1993a) if the total error probability is greater than $\alpha + 2.58s.e(\hat{\alpha})$, then the method is termed anticonservative, and if it is lower than $\alpha - 2.58s.e(\hat{\alpha})$, the method is termed conservative. The estimated error probabilities are called symmetric when the larger error probability is less than 1.5 times the smaller one.

The value of $\delta = -18$ and $\beta = 4$ were chosen as the parameters of the model to simulate failure times that mimic those seen in automobile brake pad failures. Suppose there are $i = 1, 2, \dots, n$ observations. Random numbers from the uniform distribution on the interval (0,1), u_i , was generated to produce t_i for the i^{th} observation

$$t_i = \left[\frac{1}{e^\delta} \left(\frac{1}{u_i} - 1 \right) \right]^{(1/\beta)}.$$

To obtain censored observation in our data, the censoring time for the i^{th} observation, c_i were simulated from the exponential distribution with parameter μ where the value of μ could be adjusted to obtain the desired approximate censoring proportion in our data. Following that t_i will be censored at c_i if $t_i > c_i$ and uncensored otherwise. The simulation study was carried out via the FORTRAN programming language.

Simulation results

Depicted in Table 1 are the summary of the estimated left, right and total error probabilities for the different methods discussed in the previous section. The results using the B-p method were omitted from the discussion due to the method's poor performance when compared to the other methods. The B-p interval uses the empirical distribution and tends to fail when the distribution of $\hat{\theta}^*$ is highly skewed which is rather common when bootstrapping censored and truncated observations. Inclusion of the B-p results would require a substantial increase in the graphical scale. Thus, we decided only to compare and display the results for the remaining 3 methods in all discussions that follow. More comprehensive results are given in

Tables 2–5. The left and right estimated error probabilities should preferably be equal or close to $\alpha/2$. The overall performances of the different methods were judged based on the total number of anticonservative (AC), conservative (C) and asymmetrical (AS) intervals. We are also interested in methods that behave well at different nominal levels and censoring proportions. Figures 1–8 compare the results of the coverage probability study using different methods graphically.

The alternative computer intensive methods are usually employed to relieve us from tedious calculations and asymptotic normality assumptions. Thus, we wish to see them perform well especially at smaller sample sizes where the intervals based on asymptotic normality usually fail. Based on the results of the simulation study, we see that the DB-t intervals are more reliable than the DB-p and B-t methods. The DB-t method does not produce any conservative or anticonservative intervals for both parameters δ and β , even when censoring proportion in the data is high ($cp = 15\%$).

The DB-t method produced very few asymmetrical intervals, especially for the parameter β at $\alpha = 0.05$ (see Table 1). The estimated error probabilities for the DB-t is also always closer to the nominal compared to the other methods even when the censoring proportion is high and sample size is low ($n = 25$, see Figures 1–8 and Tables 2–3). All methods seem to produce fewer conservative, anticonservative and asymmetrical intervals when $\alpha = 0.10$. So overall we can conclude that the DB-t interval is the best method to employ when dealing with censored data especially at very low sample sizes ($n = 25$). The DB-p tends to work slightly better than the B-t method but both these methods do not perform as well as the DB-t when samples sizes are low ($n = 25$).

DOUBLE BOOTSTRAP CONFIDENCE INTERVAL ESTIMATES

Table 1. Summary of the interval estimates at $\alpha = 0.05$ and 0.10 , $cp = 10\%$ and 15%

		Parameters						
		δ			β			
	Methods	C	AC	AS	C	AC	AS	
$\alpha = 0.05$	cp = 10%	Double B-p	0	1	2	0	0	4
		Single B-t	0	1	3	0	2	3
		Double B-t	0	0	1	0	0	3
	cp = 15%	Double B-p	0	1	4	0	0	4
		Single B-t	0	2	4	0	2	4
		Double B-t	0	0	1	0	0	1
$\alpha = 0.10$	cp = 10%	Double B-p	0	0	1	0	0	2
		Single B-t	0	0	2	0	0	2
		Double B-t	0	0	0	0	0	1
	cp = 15%	Double B-p	0	0	2	0	0	3
		Single B-t	0	1	3	0	1	2
		Double B-t	0	0	1	0	0	1

Table 2. Coverage probability of interval estimates for δ at $\alpha = 0.05$

		cp = 10%			cp = 15%		
Methods	n	Left	Right	Total	Left	Right	Total
Double B-p	25	0.045	0.028	0.073	0.049	0.024	0.073
	30	0.036	0.026	0.062	0.043	0.020	0.063
	40	0.030	0.022	0.052	0.033	0.020	0.053
	50	0.029	0.017	0.046	0.032	0.015	0.047
Single B-t	25	0.056	0.020	0.076	0.060	0.021	0.081
	30	0.043	0.022	0.065	0.049	0.022	0.071
	40	0.038	0.024	0.062	0.040	0.018	0.058
	50	0.037	0.025	0.062	0.040	0.022	0.062
Double B-t	25	0.029	0.028	0.057	0.033	0.024	0.057
	30	0.022	0.030	0.052	0.019	0.022	0.041
	40	0.016	0.023	0.039	0.014	0.023	0.037
	50	0.014	0.022	0.036	0.018	0.022	0.040

Table 3. Coverage probability of interval estimates for β at $\alpha = 0.05$

Methods	n	cp = 10%			cp = 15%		
		Left	Right	Total	Left	Right	Total
Double B-p	25	0.006	0.039	0.045	0.002	0.048	0.050
	30	0.008	0.033	0.041	0.004	0.038	0.042
	40	0.010	0.030	0.040	0.013	0.032	0.045
	50	0.014	0.031	0.045	0.012	0.035	0.047
Single B-t	25	0.021	0.055	0.076	0.022	0.058	0.080
	30	0.025	0.045	0.070	0.025	0.045	0.070
	40	0.022	0.035	0.057	0.023	0.043	0.066
	50	0.025	0.033	0.058	0.019	0.037	0.056
Double B-t	25	0.028	0.024	0.052	0.023	0.025	0.048
	30	0.031	0.021	0.055	0.021	0.021	0.042
	40	0.028	0.013	0.041	0.027	0.010	0.037
	50	0.027	0.015	0.042	0.022	0.015	0.037

Table 4. Coverage probability of interval estimates for δ at $\alpha = 0.10$

Methods	n	cp = 10%			cp = 15%		
		Left	Right	Total	Left	Right	Total
Double B-p	25	0.068	0.043	0.111	0.070	0.042	0.112
	30	0.057	0.041	0.098	0.064	0.041	0.105
	40	0.052	0.045	0.097	0.054	0.043	0.097
	50	0.059	0.047	0.106	0.060	0.040	0.100
Single B-t	25	0.071	0.043	0.114	0.085	0.045	0.130
	30	0.062	0.045	0.107	0.077	0.038	0.115
	40	0.063	0.044	0.107	0.070	0.045	0.115
	50	0.068	0.043	0.111	0.069	0.047	0.116
Double B-t	25	0.043	0.053	0.096	0.049	0.043	0.092
	30	0.037	0.053	0.090	0.033	0.045	0.078
	40	0.040	0.047	0.087	0.030	0.050	0.080
	50	0.041	0.046	0.087	0.050	0.046	0.096

DOUBLE BOOTSTRAP CONFIDENCE INTERVAL ESTIMATES

Table 5. Coverage probability of interval estimates for β at $\alpha = 0.10$

Methods	n	cp = 10%			cp = 15%		
		Left	Right	Total	Left	Right	Total
Double B-p	25	0.036	0.064	0.100	0.029	0.071	0.100
	30	0.037	0.059	0.096	0.035	0.068	0.103
	40	0.040	0.052	0.092	0.038	0.054	0.092
	50	0.040	0.056	0.096	0.036	0.055	0.091
Single B-t	25	0.046	0.077	0.123	0.046	0.086	0.132
	30	0.046	0.065	0.111	0.040	0.078	0.118
	40	0.040	0.069	0.109	0.047	0.066	0.113
	50	0.046	0.065	0.111	0.049	0.066	0.115
Double B-t	25	0.054	0.040	0.094	0.051	0.042	0.093
	30	0.055	0.035	0.090	0.047	0.032	0.079
	40	0.048	0.032	0.080	0.050	0.027	0.077
	50	0.048	0.039	0.087	0.046	0.039	0.085

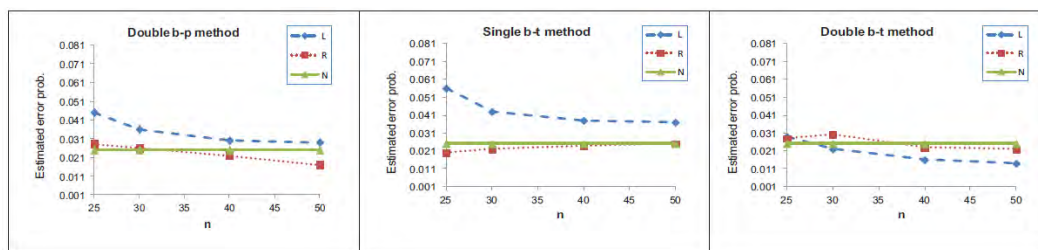


Figure 1. Interval estimates at $\alpha = 0.05$, $cp = 10\%$ for parameter δ

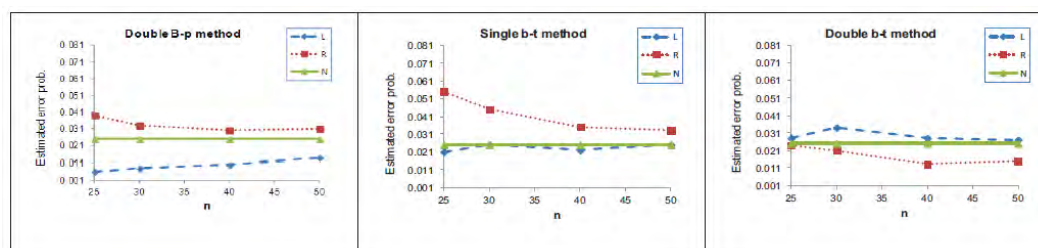


Figure 2. Interval estimates at $\alpha = 0.05$, $cp = 10\%$ for parameter β

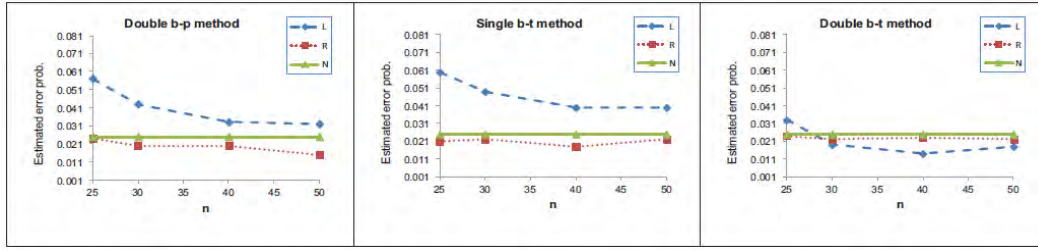


Figure 3. Interval estimates at $\alpha = 0.05$, $cp = 15\%$ for parameter δ

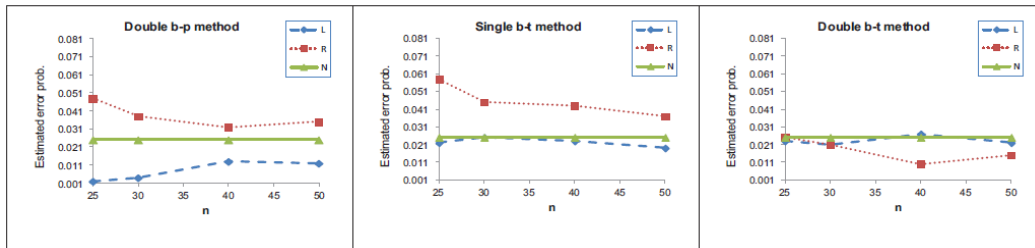


Figure 4. Interval estimates at $\alpha = 0.05$, $cp = 15\%$ for parameter β

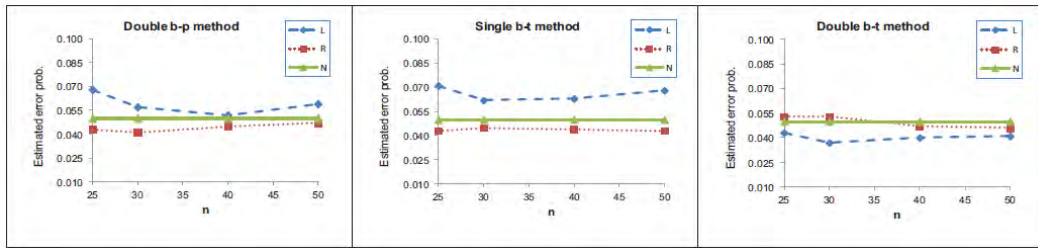


Figure 5. Interval estimates at $\alpha = 0.10$, $cp = 10\%$ for parameter δ

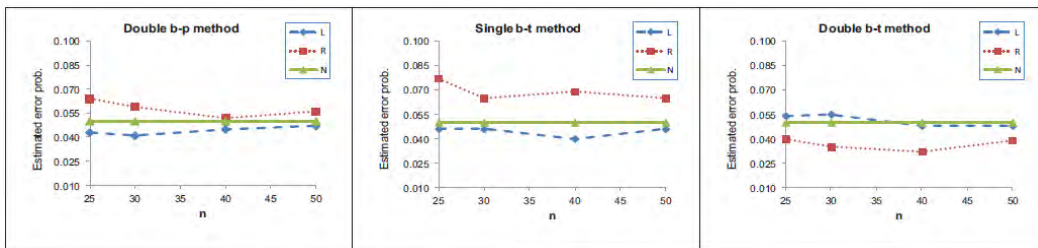


Figure 6. Interval estimates at $\alpha = 0.10$, $cp = 10\%$ for parameter β

DOUBLE BOOTSTRAP CONFIDENCE INTERVAL ESTIMATES

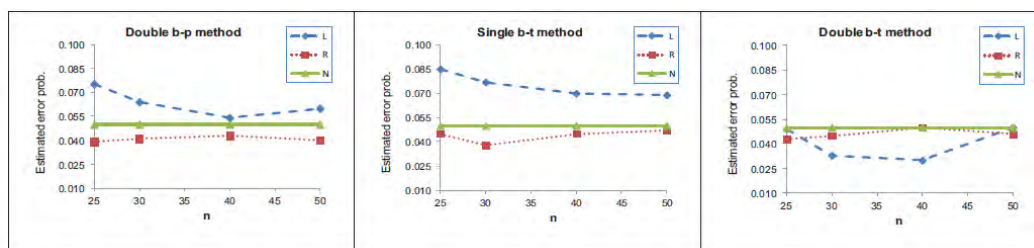


Figure 7. Interval estimates at $\alpha = 0.10$, $cp = 15\%$ for parameter δ

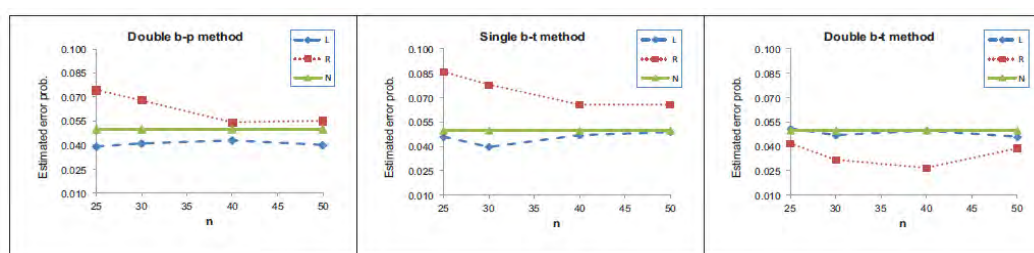


Figure 8. Interval estimates at $\alpha = 0.10$, $cp = 15\%$ for parameter β

Real data analysis

To illustrate the application of the double bootstrap confidence interval procedures, we will consider two data sets dealing with censored and truncated observations respectively. First is the data on breast cancer by Leathem and Brooks (1987) on the lifetimes of 32 women whose tumor has potential to metastasize thus classified as positive staining. 11 of the observations were censored which make the censoring proportion almost 34%.

Considering the sample size and the censoring proportion in the data, any use of inferential procedures based on the asymptotic normality of the maximum likelihood estimates is not advisable. The second is 98 left truncated data on the lifetimes of the brake pads of automobiles. The left truncated lifetimes u_i is the current odometer reading for each car. Only cars that had initial pads were selected and the remaining, and initial pad thickness, were used to estimate t_i . Although the simulation study did not extend to include any truncated data, we believe we may generalize the results obtained using censored data to truncated data since in both cases the distribution of the bootstrap estimates will be skewed and far from normal, see Figures 10 and 12.

Right censored data

The log logistic distribution fits the breast cancer data well as shown by the probability plot in Figure 9. Table 6 shows the parameter estimates when the log logistic distribution is fit to the data and the 95% confidence intervals using the Wald, DB-t, DB-p and B-t interval estimation procedures.

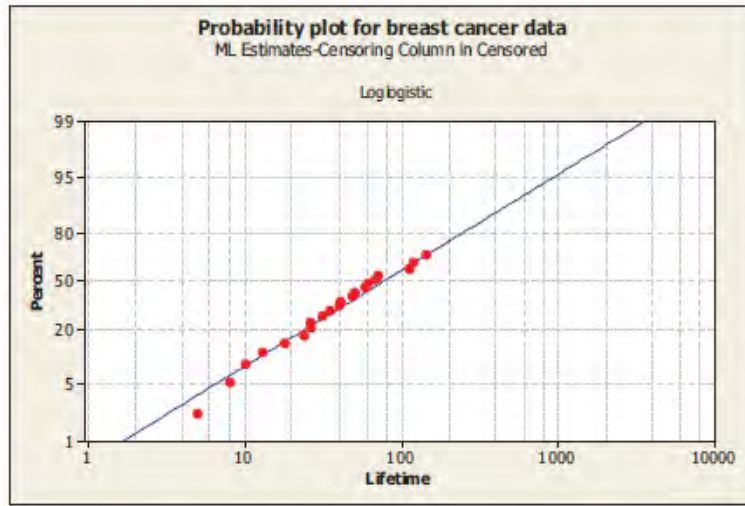


Figure 9. Log logistic probability plot for the breast cancer data

Table 6. MLE of breast cancer data and 95% confidence intervals

θ	$\hat{\theta}$	SE	Wald	B-t	DB-p	DB-t
δ	-5.187	0.953	(-7.054,-3.3189)	(-7.154,-3.549)	(-6.824,-3.744)	(-8.231,-3.824)
β	1.200	0.220	(0.769,1.633)	(0.811,1.666)	(0.826,1.611)	(0.824,1.677)

Figure 10 shows the histogram of 1000 bootstrap replications of $\hat{\delta}$ and $\hat{\beta}$. We can clearly see that both the histogram are not very close to normal shape and appear to be skewed especially the distribution of $\hat{\delta}^*$. Goodness of fit test based on the Anderson-Darling and Kolmogorov-Smirnov had also rejected the assumption of normality at $\alpha = 0.05$. In this case we can expect a disagreement between the standard normal interval and intervals based on the bootstrap methods as highlighted by Efron and Tibshirani (1993).

DOUBLE BOOTSTRAP CONFIDENCE INTERVAL ESTIMATES

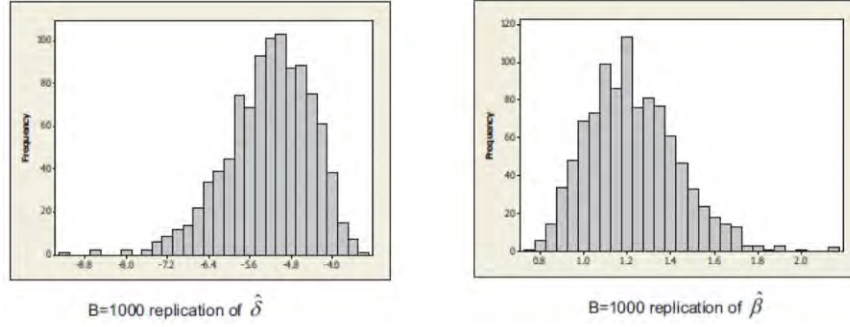


Figure 10. Bootstrap replication of $\hat{\delta}$ and $\hat{\beta}$ for breast cancer data

Based on the coverage probability study, we could see the DB-t intervals produces the most reliable intervals when sample sizes are small and data is highly censored. From Figure 10, we can see that the distribution of $\hat{\delta}^*$ is more skewed than $\hat{\beta}^*$. Thus, as expected the DB-t interval for the parameter δ tends to disagree more with the Wald interval, than the parameter β . The DB-t interval for the parameter δ is also wider than other intervals because it tries to accommodate the skewness in the distribution of the bootstrap estimates, which eventually increases the probability of the true parameter value lying within this interval as verified by the results of the coverage probability study. Other intervals, though narrower may fail to include the true parameter value. Observe that for the parameter β the Wald interval is wider than the DB-t intervals. So the DB-t interval has the ability to adjust itself according to the distribution of the bootstrap estimates which is directly linked to the data in hand and not dependent on any theoretical assumptions, which may fail when the normal approximation is simply not true. So it would actually be more practical to employ the DB-t interval in this case.

Based on the results of the coverage probability study we can see that the DB-p and B-t intervals do not perform as well as the DB-t when sample sizes are small. Their performance tends to improve gradually when $n > 30$. However we have included the interval estimates based on these methods merely to do some comparison study. As we can see the B-t interval is the narrowest among the 4 intervals and there is much doubt as to whether the true parameter value will actually be included within this interval. The DB-p interval is much closer to the Wald interval than the DB-t, especially for parameter δ . This only makes it clearer that the DB-t interval will be more reliable since DB-p did not perform as well as DB-t in the coverage probability study.

Left truncated data

Table 7: MLE of brake pad data and 95% confidence intervals

θ	$\hat{\theta}$	SE	Wald	B-t	DB-p	DB-t
δ	-17.062	1.619	(-20.235,-13.889)	(-20.381,-13.810)	(-20.246,-13.765)	(-20.686,-14.005)
β	4.137	0.383	(3.387,4.888)	(3.397,4.892)	(3.379,4.831)	(3.427,4.974)

Table 7 shows the parameter estimates when the log logistic distribution is fit to the brake pad data and the 95% confidence intervals using the Wald, DB-t, DB-p and

B-t interval estimation procedures. It is known that $e_i = \frac{S(t_i; \hat{\theta})}{S(u_i; \hat{\theta})}$ is $U(0,1)$ given u_i .

Thus, can plot the uniform residual, e_i against the uniform quantile to see if the log logistic distribution fits data as given in Figure 11. As we can see the model fits the data quite well.

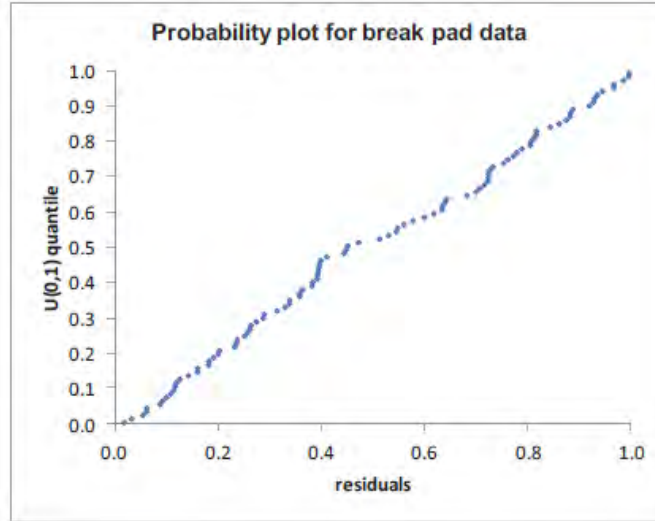


Figure 11. Log logistic probability plot for break pad data

Figure 12 shows the histogram of 1000 bootstrap replications of $\hat{\delta}$ and $\hat{\beta}$. We can also observe that for the left truncated data both the histogram appear to be skewed. The goodness of fit test based on the Anderson-Darling and Kolmogorov-

DOUBLE BOOTSTRAP CONFIDENCE INTERVAL ESTIMATES

Smirnov had again rejected the assumption of normality at $\alpha = 0.05$. As with the breast cancer data we can expect a disagreement between the standard normal interval and intervals based on the bootstrap methods.

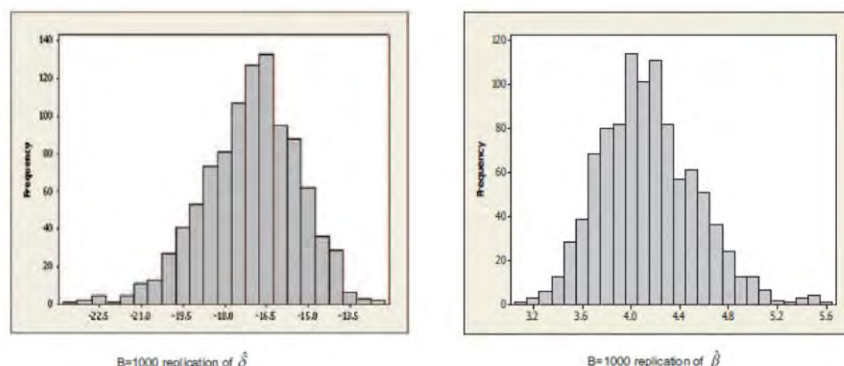


Figure 12. Bootstrap replication of $\hat{\delta}$ and $\hat{\beta}$ for break pad data

The DB-t interval for both the parameters tends to disagree with the Wald interval due to the skewness in the distribution of the bootstrap estimates. The DB-p gives the narrowest interval followed by the Wald. However as we discussed earlier we should be cautious since these intervals may fail to include the true parameter value. Similar pattern is displayed for parameter β where the DB-t interval is the widest but now the DB-p interval is the narrowest. For both parameters the B-t interval and DB-p interval seem to be rather close to the Wald interval which again makes it clearer that the DB-t interval will be more reliable as the B-t and DB-p interval did not perform as well as the DB-t in the coverage probability study.

Discussion

It may appear as if the Wald would suffice as a confidence interval estimate due to its simplicity but this may not be true with smaller data sets that are censored or truncated. So while the Wald can still be employed especially when samples sizes are large and censoring proportions in the data is low, alternative bootstrap methods such as the DB-t should be employed otherwise. Although the DB-t method is slightly more computational compared to the Wald, with the existence of fast computers and parallel computing techniques, these results can be obtained very

quickly especially for small data sets. The performances of the B-t and DB-p method do not seem to be significantly better than the DB-t methods. Thus, we would not recommend employing them for the construction of confidence intervals for the parameters of this model.

Keep in mind that the bootstrap confidence interval methods was initially introduced by Efron and Tibshirani (1993) mostly for use with location statistics such as the mean, median and trimmed mean. However, recently it is being used for more complicated statistics, especially those with standard errors that are attainable. Through simulation studies, we are able to assess if the bootstrap methods can be extended to other general problems, such as constructing confidence interval estimates for model parameters with censored and truncated data. They offer an alternative rather than depending solely on interval estimates that are based on asymptotic normality theory.

Although many are still skeptical about these methods, we can't deny that they provide us with an opportunity to perform comparison study which in some cases may lead to estimates that are better than those produced by traditional methods. The computation time for the double bootstrap is 2 times longer than the ordinary or single bootstrap procedures and negligible for the Wald procedure. For analyzing the breast cancer data, the computation time using the FORTRAN programming language only took 0.015 seconds for the single bootstrap procedures and 0.031 seconds for double bootstrap procedures. The single bootstrap procedure for the brake pad data took 0.03 seconds and the double bootstrap procedures took 0.06 seconds. So, the results can be obtained extremely quickly in only matter of seconds. Thus, the argument of bootstrap methods being heavily computational is not applicable anymore due to the availability of very fast computers.

The methods discussed here can be applied to the parameters of other survival models involving censored or truncated data. The log logistic model discussed here was chosen mainly due to its popularity in most cancer studies and its ability to accommodate both fixed and time dependent covariates easily. When the data in hand has more complicated structures such as double or interval censored the construction of the confidence intervals using solely methods based on asymptotic normality becomes unreliable. Thus, it's important and also interesting to see if these alternative methods provide us with more appealing solutions.

References

- Arasan, J., & Lunn, M. (2008). Alternative interval estimation for parameters of bivariate exponential model with time varying covariate. *Computational Statistics*, 23, 605–622.
- Arasan, J., & Lunn, M. (2009). Survival model of a parallel system with dependent failures and time varying covariates. *Journal of Statistical Planning and Inference*, 139(3), 944–951.
- Bennet, D. S. (1983). Log-Logistic Regression Models for Survival Data. *Applied Statistics*, 32(2), 165–171.
- Cox, D. R., & Lewis, P. A. (1966). *Statistical Analysis of Series of Events*. London: Methuen.
- Cox, D. R., & Oakes, D. (1984). *Analysis of Survival Data*. New York: Chapman and Hall.
- Diciccio, T. J., & Efron, B. (1996). Bootstrap confidence interval. *Statistical Science*, 11(3), 189–228.
- Doganaksoy, N. & Schmee, J. (1993a). Comparison of approximate confidence intervals for distributions used in life-data analysis. *Technometrics*, 35(2), 175–184.
- Doganaksoy, N., & Schmee, J. (1993b). Comparison of approximate confidence intervals for smallest extreme value distribution simple linear regression model under time censoring. *Communications in Statistics – Simulation and Computation*, 2, 175–184.
- Efron, B. (1981a). Nonparametric standard errors and confidence intervals. *Canadian Journal of Statistics*, 9, 139–172.
- Efron, B. (1981b). Nonparametric standard errors: the jackknife, the bootstrap and other methods. *Biometrika*, 68, 589–599.
- Efron, B. (1985). Bootstrap confidence interval for a class of parametric problem. *Biometrika*, 72, 45–58.
- Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman and Hall.
- Gupta, R. C., Akman, O., & Lvin, S. (1999). A study of log-logistic model in survival analysis. *Biometrical Journal*, 41(4), 431–443.
- Hall, P. (1986). On the bootstrap and confidence intervals. *Annals of Statistics*, 14, 1431–1452.

- Hall, P. (1988a). Theoretical comparison of bootstrap confidence intervals (with discussion). *Annals of Statistics*, 16, 927–953.
- Hall, P. (1988b). On symmetric bootstrap confidence intervals. *Journal of Royal Statistical Society, B*(50), 35–45.
- Jeng, S. L., & Meeker, W. Q. (2000). Comparison for approximate confidence interval procedure for type I censored data. *Technometrics*, 42(2), 135–148.
- Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*. New York: John Wiley and Sons.
- Leathem, A. J., & Brooks, S. (1987). Predictive value of lectin binding on breast-cancer recurrence and survival. *The Lancet*, 329(8541), 1054-1056.
- Letson, D. & McCullough, B. D. (1998). Better confidence intervals: the double bootstrap with no pivot. *American journal of agricultural economics*, 80 (3), 552–559.
- Martin, M. A. (1992). On the double bootstrap. New York: Springer.
- Mazucheli, J., Barros, E. A. C., & Achcar, J. A. (2005). Bootstrap confidence intervals for the mode of the hazard function. *Computer Methods and Programs in Biomedicine*, 79(1), 39–47.
- McCullough, B. D., & Vinod, H. D. (1998). Implementing the double bootstrap. *Computational Economics*, 12(1), 79–95.
- O’Quigley, J. & Struthers, L. (1982). Survival models based upon the logistic and log-logistic distributions. *Computer Programs in Biomedicine*, 15(1), 3–11.
- Robinson, J. A. (1983). Bootstrap confidence intervals in location-scale models with progressive censoring. *Technometrics*, 25(2), 179–187.
- Schenker, N. (1985). Qualms about bootstrap confidence intervals. *Journal of American Statistical Association*, 80, 360–361.
- Shi, S. G. (1992). Accurate and efficient double-bootstrap confidence limit method. *Computational Statistics & Data Analysis*, 13(1), 21–32.
- Singh, K. (1981). On the asymptotic accuracy of Efron’s bootstrap. *Annals of Statistics*, 9, 1187–1195.

Contrast of Bayesian and Classical Sample Size Determination

Farhana Sadia

University of Dhaka
Dhaka, Bangladesh

Syed S. Hossain

University of Dhaka
Dhaka, Bangladesh

Sample size determination is a prerequisite for statistical surveys. A comprehensive overview of the Bayesian approach for computation of the sample size, and a comparison with classical approaches, is presented. Two surveys are taken as example to illustrate the accuracy and efficiency of each approach, and to make recommendations about which method is preferred. The Bayesian approach of sample size determination may require fewer subjects if proper prior information is available.

Keywords: Sample size determination, Bayesian methods, mean

Introduction

A good statistical study is one that is well designed and leads to a valid conclusion. The main aim of the sample size determination (SSD) is to find an adequate number of observations to be made to estimate the population prevalence with a good precision. That means an optimal sample size is required to give a desired level of validity of the results. Prior determination of a good sample size reduces expenses and time by allowing researchers to estimate information about a whole population without having to survey each member of the population (Cochran, 1977). A considerable number of criteria for SSD are available depending on the two types of inferential approaches-Frequentist and Bayesian.

Frequentist sample size determination methods depend directly on the unknown parameter of interest which in practice is often very hard to get whereas Bayesian way does not depend on the guessed value of the true parameter rather it depends on a prior distribution of the parameter (M'lan et al., 2008). Bayesian methods often results in providing a posterior distribution which combines the pre-experimental information of the parameter (prior distribution) with the

Pr. Sadia is a Lecturer in the Institute of Statistical Research and Training (ISRT). Email him at: fsadia@isrt.ac.bd. Dr. Hossain is Professor in the Institute of Statistical Research and Training (ISRT). Email him at: shahadat@isrt.ac.bd.

experimental data by utilizing the likelihood of the parameter (Pham-Gia, 1995). In case of Bayesian sample size determination, the marginal or prior-predictive distribution is used which is the mixture of the sampling distribution of the data and the prior distribution of the unknown parameters (M'lan et al., 2008). In this context, the minimal sample size determination using three different Bayesian approaches based on highest posterior density (HPD) intervals which are average coverage criterion (ACC), average length criterion (ALC) and worst outcome criterion (WOC) (Joseph et al., 1995) are examined herein. Sample sizes for two real life surveys were calculated using these criteria and were compared with the sample size determined by classical method as well as with the actual sample size utilized in these surveys.

Methodology

Bayesian Methods Used in Sample Size Determination

Let θ be an unknown parameter vector that is derived to be estimated and Θ be the parameter space for θ . Suppose it is desired to determine the sample size n where a random sample $X = (X_1, X_2, \dots, X_n)$ is to be used for the estimation of θ . According to the Bayesian approach, if $f(\theta)$ is the prior distribution for the parameter and the likelihood function given the data $x = (x_1, x_2, \dots, x_n)$ is $L(\theta; x) \propto f(x|\theta)$. The preposterior marginal distribution of x is thus given by

$$f(x) = \int_{\Theta} f(x|\theta) f(\theta) d\theta. \quad (1)$$

Now, the posterior distribution of θ given data x with sample size n is

$$f(\theta|x, n) = \frac{f(x|\theta) f(\theta)}{\int_{\Theta} f(x|\theta) f(\theta) d\theta}. \quad (2)$$

Using HPD interval approach, a sample size n most appropriate for estimating θ can be obtained by finding the n that gives the highest coverage of the equation (2) for a given fixed interval. The following three criteria are used in this paper. For details of these criteria, see (M'lan et al., 2008; Joseph et al., 1995; Joseph et al., 1997; Sahu et al., 2006). The average coverage criterion seeks the smallest n satisfying the following condition

CONTRAST OF TWO SAMPLE SIZE DETERMINATIONS

$$\int_{\mathcal{X}} \left\{ \int_{a(x,n)}^{a(x,n)+l} f(\theta | x, n) d\theta \right\} f(x) dx \geq 1 - \alpha, \quad (3)$$

where, $f(x)$ and $f(\theta | x, n)$ are given in equation (1) and (2) and $a(x, n)$ is the lower limit of the HPD credible set of length l for the posterior density $f(\theta | x, n)$. In general, $a(x, n)$ will depend both on the data x and the sample size n . ACC finds the minimum sample size n such that the expected coverage probability is at least $(1 - \alpha)$ for a given fixed HPD interval length l . The average length criterion seeks the smallest n satisfying the condition

$$\int_{\mathcal{X}} l'(x, n) f(x) dx \leq l, \quad (4)$$

where l is the desired pre-specified average length. This average length criterion is used to find a sample size n that would fix the coverage probability $(1 - \alpha)$ of the HPD credible set for θ . The worst outcome criterion finds the smallest n satisfying the following condition

$$\inf_{x \in \mathcal{X}} \left\{ \int_{a(x,n)}^{a(x,n)+l} f(\theta | x, n) d\theta \right\} \geq 1 - \alpha, \quad (5)$$

where both l and α are fixed in advance.

Bayesian Sample Sizes for Normal Mean

Let the data vector $x = (x_1, x_2, \dots, x_n)$ consist of exchangeable components from a normal distribution with the unknown normal mean μ and variance σ^2 . The precision of the data is then $\lambda = \sigma^2$. In this case, the prior distribution is a conjugate prior distribution. The prior distribution for μ and λ are $\lambda \sim \text{gamma}(\nu, \beta)$ and $\mu | \lambda \sim N(\mu_0, n_0 \lambda)$. That means, the conjugate prior distribution for (μ, λ) is the normal-gamma conjugate prior distribution.

Sample Sizes for Single Normal Mean with Known Precision

If the precision λ is known, then the posterior distribution will be a normal distribution, i.e.

$$\mu | x \sim N(\mu_n, \lambda_n), \quad (6)$$

where

$$\begin{aligned}\lambda_n &= (n + n_0) \lambda, \\ \mu_n &= \frac{n_0 \mu_0 + n \bar{x}}{n_0 + n}, \\ \text{and } \bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i\end{aligned}$$

In this case, the three Bayesian sample size criteria give the same solution because the posterior precision depends only on n and does not vary with the particular observed data vector x . This is also equivalent to that given by Adcock (1988) as

$$n \geq \frac{4Z_{1-\alpha/2}^2}{\lambda^2} - n_0. \quad (7)$$

If a non-informative prior is used such that $n_0 = 0$, then inequality (7) reduces to the classical formulation.

Sample Sizes for Single Normal Mean with Unknown Precision

If the precision λ is unknown, then marginal posterior distribution of λ is given by

$$\mu | x \sim \sqrt{\left\{ \frac{\beta_k}{(n + n_0)(v + \frac{n}{2})} \right\}} t_{2v+n} \mu_n, \quad (8)$$

where

$$\beta_n = \beta + \frac{1}{2} n s^2 + \frac{1}{2} \frac{n n_0}{n + n_0} (\bar{x} - \mu_0)^2,$$

and t_{2v+n} represents a t-distribution with $2v + n$ degrees of freedom. In this case, different Bayesian sample size criteria will give different sample size if the posterior precision varies with the data.

The ACC sample size for unknown precision is similar to that for known precision because $v | \beta$ is the prior mean for precision λ , thus it is only necessary to substitute the prior mean precision for λ in inequality (7) and exchange the normal

CONTRAST OF TWO SAMPLE SIZE DETERMINATIONS

quantile Z with a quantile from a t_{2v} distribution. If the degrees of freedom of t distribution do not increase with the sample size, equation (7) can give different sample size which is substantially different from those from inequality (8) and classical method of sample size.

The average length criterion seeks the minimum n satisfying the following condition

$$2t_{n+2v, 1-\alpha/2} \sqrt{\left\{ \frac{2\beta}{(n+2v)(n+n_0)} \right\} \frac{\Gamma\left(\frac{n+2v}{2}\right)\Gamma\left(\frac{2v-1}{2}\right)}{\Gamma\left(\frac{n+2v-1}{2}\right)\Gamma(v)}} \leq l. \quad (9)$$

When estimating a single normal mean with unknown precision λ with a gamma (v, β) prior distribution on λ , the ALC (4) is satisfied for large n . Although it does not appear feasible to solve inequality (9) explicitly for n , the left-hand side is straight forward to calculate given v, β, α and n . Therefore, the exact smallest n can be found by a bi-sectional search algorithm (Chen et al., 1998).

For a single normal mean with unknown precision λ with a gamma (v, β) prior distribution on λ , the WOC is satisfied when n is sufficiently large so that

$$\frac{l^2 (n+2v)(n+n_0)}{8\beta \left\{ 1 + \left(\frac{n}{2v} \right) F_{n, 2v, 1-w} \right\}} \geq t_{n+2v, 1-\alpha/2}^2, \quad (10)$$

where $F_{n, 2v, 1-w}$ denotes the $100(1-w)$ percentile of an F distribution with n and $2v$ degrees of freedom and $\int_{\chi} f(x) dx = 1-w$. Therefore, the exact smallest n satisfying inequality (10) can be found by a bi-sectional search algorithm. If $X = \chi$, the sample size is not defined, because $F_{n, 2v, 1-w} \rightarrow \infty$ as $w \rightarrow 0$, hence inequality (10) cannot be satisfied for any n .

Results

Classical and Bayesian Sample Size for mean with Simple Random Sampling

For simple random sampling, computation of classical sample size for mean is made using the conventional formula (Cochran, 1977)

$$n = \frac{z_{\frac{\alpha}{2}}^2 (CV)^2}{r^2}, \quad (11)$$

where, CV is the coefficient of variation and r the relative margin of error. Also note that with the population mean denoted by μ , $d = r\mu$ where d is absolute the margin of error and α is the level of significance. Pre-assigned values of α , r , CV can give an appropriate sample size n .

Bayesian sample sizes are computed using the three types of criteria given in earlier section and these criteria find the minimum sample size n satisfying the respective condition of the criteria. Table 1 gives the classical and Bayesian sample size for mean with $\alpha = 0.05$ considering simple random sampling assuming different prior distributions of mean. Also note that different length of the posterior credible interval for the mean are computed and given in Table 1. To make sense of Bayesian sample size in Table 1, gamma prior distributions for the precision are used with different types of parameters.

In Table 1 the coefficient of variation used in the classical method is $CV=2$. Table 1 shows that the three Bayesian criteria provides different sample sizes. It is also observed from Table 1 that Bayesian criteria ACC and WOC seem to lead similar sample sizes whereas ALC criteria provides the smallest sample sizes. For example, in case of $l = 0.1$ and a prior about mean, $u = v = 2$, ACC and WOC yield the sample size of $n = 3074$ and $n = 3638$ which are somewhat similar but ALC yields a sample size of $n = 2405$ which is smaller than that using ACC and WOC. However, from Table 1 the theoretical knowledge that $n_{ALC} \leq n_{ACC} \leq n_{WOC}$ is observed to be satisfied. It is important to note that, as long as non-informative prior approaches to informative prior, the sample size gradually reduces. For example, Bayesian sample sizes are larger than classical sample size for non-informative prior (1, 1) but they are smaller than the classical sample size when using the more informative prior. That means, if more informative prior information is in hand, Bayesian method could supply more parsimonious sample size than classical method would.

CONTRAST OF TWO SAMPLE SIZE DETERMINATIONS

Table 1. Classical and Bayesian Sample Size for SRS ($\alpha = 0.05$)

Length (l)	Classical sample size	Different prior	Bayesian sample size		
			ACC	ALC	WOC
0.1	6147	Gamma (1,1)	7396	4819	7948
		Gamma (2,2)	3074	2405	3638
		Gamma (3,3)	2385	2028	2627
		Gamma (4,4)	2118	1877	2488
		Gamma (2,3)	4526	3612	4962
0.2	1537	Gamma (1,1)	1842	1198	2010
		Gamma (2,2)	761	595	823
		Gamma (3,3)	589	501	675
		Gamma (4,4)	522	463	614
		Gamma (2,3)	1147	463	1250
0.3	683	Gamma (1,1)	813	528	890
		Gamma (2,2)	333	260	392
		Gamma (3,3)	257	218	310
		Gamma (4,4)	227	201	255
		Gamma (2,3)	504	394	560
0.5	246	Gamma (1,1)	287	185	311
		Gamma (2,2)	114	88	136
		Gamma (3,3)	86	73	105
		Gamma (4,4)	76	67	89
		Gamma (2,3)	176	136	190

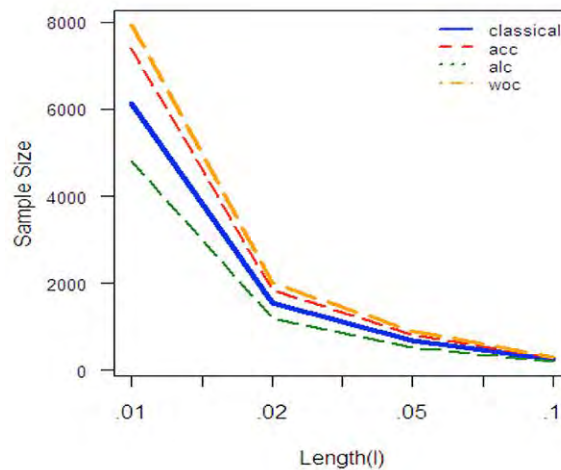


Figure 1. Classical and Bayesian sample size for different length with prior gamma (1,1)

Figure 1 gives the the comparison among classical sample sizes and Bayesian sample sizes for mean with respect to different length of the highest posterior density interval using the non-informative prior (1,1). Figure 1 shows that the Bayesian WOC criteria provides the largest sample size whereas Bayesian ALC criteria provides the smallest sample size and classical sample size and Bayesian ACC criteria give almost similar sample size. Figure 1 also elucidates that as length increases, sample size gradually decreases and this fact is true for both classical and Bayesian method of sample size determination.

Applicability of the Bayesian SSD in Real Life

The sample size determination (SSD) approaches from Bayesian perspective are grounded in theory and are eventually candidates for utilization in some real surveys. However, positive utilization of these methods in large-scale survey research in Bangladesh would depend on the computational features of the methods with respect to those usually used in such surveys. This study considered two recently conducted surveys as examples by comparing the sample sizes in these surveys with the hypothetical appropriate sample size computed using Bayesian criteria. The choice of the surveys was arbitrary; samples were selected mainly by considering availability in published format.

Household Income and Expenditure Survey 2010

Household Income and Expenditure Survey (HIES) is conducted by Bangladesh Bureau of Statistics (BBS), and is the main data source for estimation of poverty in Bangladesh. This survey provides valuable data on household income, expenditure, consumption, savings, housing condition, education, employment, health and sanitation, water supply and electricity, etc. (HIES, 2005). In the 2010 survey, a two-stage stratified random sampling technique was followed in drawing samples. The sample size of HIES 2010 was reported as 12,240 households, where 7,840 were from rural areas and 4,400 from urban areas. For making theoretically comparable, the required sample size was also calculated using the usual classical formula in equation (11) and multiplying it by design effect (deff) for adjustment of cluster sampling. Note that the choice of design effect = 1.6 is made on the basis of conventional practice in Bangladesh surveys where design effect is assumed to vary from 1.5 to 2.0.

In this computation $CV(x)$, the pre-assumed value of the population coefficient of variation is computed from the HIES 2005 considering “Household Income” as the main interesting variable, $z^2_{\frac{\alpha}{2}} = 1.64$ and the maximum allowable

CONTRAST OF TWO SAMPLE SIZE DETERMINATIONS

relative margin of error $r = 10\%$. The sample sizes actually used in HIES 2010 along with sample sizes computed using classical and Bayesian methods are given in Table 2. For the Bayesian approach the prior $(\alpha, \beta) = \left(\frac{1}{\sqrt{CV}}, std\right)$ is considered and the CV and standard deviation of “Household Income” computed from the HIES 2005 are used. The CV for rural and urban populations as 3.01 and 4.71 respectively are used in both approaches of sample size determination.

Table 2. Classical and Bayesian Sample Size for HIES 2010 ($\alpha = 0.1$)

Region	Sample size actually used	Classical method	Hypothetically computed sample size		
			ACC	ALC	WOC
Rural	7840	3922	3621	8542	7096
Urban	4400	9604	9200	4908	5021
Total	12240	13526	12821	13450	12117

From Table 2, it can be observed that the total sample size used in the actual survey is almost same as that determined by the classical method as well as by the three Bayesian criteria. However, the urban-rural split of the sample sizes seems to be of reverse order in the hypothetically determined methods. This could be due to the reason that the actual study allocated the size proportionally to the 70%-30% rural-urban population of Bangladesh whereas the classical and Bayesian SSD used in the hypothetical computation considered separate sample sizes for urban and rural domains, and because CV of household income in urban area is much higher than that in rural area, the urban sample sizes are obtained to be larger. It is obvious that the choice of higher sample size in urban area according to the computed sample size could have provided better precision than that possibly been attained in the actual survey.

The comparison between the classical and Bayesian SSD for the said survey reveals not much of difference except that the WOC criteria produced smallest sample size in comparison to the other methods. The ACC criteria and the classical method give almost a same sample size, which implies that with similar level of prior information even the most conservative Bayesian criteria gives as good sample size as the classical method. This result has been revealed in an extensive simulation with different level of significance and different level of precision. However, it can be expected that if higher level of prior information is in hand, Bayesian approach may possibly utilize them and reduce the required number of

samples whereas the classical method do not have any option to utilize them. That means that, if a Bayesian approach is applied, as opposed to a classical approach for sample size determination, then it could have optimized the opportunity.

Bangladesh Demographic and Health Survey (BDHS) 2007

The Bangladesh Demographic and Health Survey (BDHS) is a periodic survey conducted in Bangladesh to serve as a source of population and health data for policymakers, program managers, and the research community. The sample size for [BDHS 2007](#) was determined according to six divisions and two regions using [BDHS 2004](#) with the help of the usual SSD formula (see the previous section) and the three Bayesian criteria (as described in the Methodology).

Table 3. Classical and Bayesian Sample size for six divisions and two regions of BDHS 2007 ($\alpha = 0.05$)

Region	Sample size actually used	Classical method	Hypothetically computed sample size		
			ACC	ALC	WOC
Dhaka	2726	376	394	213	398
Chittagong	2423	448	468	259	456
Khulna	1935	683	1124	677	1010
Barisal	1674	1071	1490	912	1256
Rajshahi	2403	707	1064	637	955
Sylhet	1949	267	166	69	187
Total (for division)	11485	3552	4706	2767	4262
Urban	5218	690	1111	669	998
Rural	7981	513	605	346	576
Total (for region)	11485	1203	1716	1015	1574

Considering the variable “children ever born” as the variable of interest, computations similar to those in the previous section were done. The actual sample sizes used in the survey along with the computed required sample sizes using Bayesian and classical methods are given in [Table 3](#). The sample sizes are computed with two different perspectives about domains. Often only the rural/urban segregation of the estimates is needed from surveys; in such cases sample size may be calculated for only those two domains. BDHS 2007 makes separate estimates for the six administrative divisions of Bangladesh and hence these six domains were considered in the computation.

CONTRAST OF TWO SAMPLE SIZE DETERMINATIONS

Table 3 shows that the hypothetically determined sample size (classical and Bayesian approach) is very much smaller than actually used sample size of BDHS 2007 for division and regions which indicates oversampling for reporting more reliable estimates of the rarer characteristics of that division or region. However, classical and Bayesian sample sizes are determined using the usual SSD formula due to unavailability of the used sample size formula of BDHS 2007. Because the coefficient of variation of “children ever born” is very low, so that a very much smaller sample size was obtained from classical and Bayesian approach than the used sample size of BDHS 2007. This may be explained because the Bayesian sample size using ACC and WOC criteria is larger than the classical sample size for all division and two regions. This table concludes that the Sylhet division has smallest sample size among all divisions for both approach and actually used sample size of BDHS 2007. Also note that Barisal division has the largest sample size among the other divisions of Bangladesh for classical and Bayesian approach but BDHS 2007 showed that the Dhaka division has the largest sample size among all divisions. This table also shows that the urban-rural sample size is present in reverse order in the hypothetically determined methods like the previous survey (HIES 2010). This statement indicates that the sample size allocation among the urban-rural strata and among the divisions could have possibly been done proportionally.

Conclusion

Results suggest that the classical sample size is larger than Bayesian sample size in the applications examined, although the estimated sample sizes in both methods (classical and Bayesian) are decreased when a larger margin of error is considered. Prior information can reasonably be utilized to improve Bayesian sample size estimation. In Bayesian approach of sample size determination, different prior are used in place of classical estimator. The estimated sample sizes decreased when moving towards informative prior from a non-informative prior. Results from this study show that the proper use of prior information may enhance the strength of the Bayesian method of sample size determination. Thus, an optimized parsimony could be achieved by use of Bayesian sample size determination with substantially informative priors.

References

- Adcock, C. J. (1988). A Bayesian approach to calculating sample sizes. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 37, 433-439.
- BDHS (Bangladesh Demographic and Health Survey) 2004, National Institute of Population Research and Training (NIPORT), Mitra and Associates and ORC Macro, 2005, NIPORT, Dhaka.
- BDHS (Bangladesh Demographic and Health Survey) 2007, National Institute of Population Research and Training (NIPORT), Mitra and Associates and ORC Macro, 2008, NIPORT, Dhaka.
- Chen, M. H., & Shao, Q. M. (1998). Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics*, 7.
- Cochran, W. G. (1977). *Sampling Techniques* (3rd ed.). New York: John Wiley and Sons.
- Joseph, L., & Belisle, P. (1997). Bayesian sample size determination for normal means and difference between normal means. *Royal Statistical Society*, 46(2), 209-226.
- Joseph, L., Wolfson, D. B., & Berger, R. D. U. (1995). Sample size calculations for binomial proportions via highest posterior density intervals. *Royal Statistical Society*, 44(2), 143-154.
- M'LAN, C. E., Joseph, L., & Wolfson, D. B. (2008). Bayesian sample size determination for binomial proportions. *International Society for Bayesian Analysis*, 3(2), 269-296.
- Pham-Gia, T. (1995). Sample size determination in Bayesian statistics: A commentary. *Journal of the Royal Statistical Society*, 44(2), 163-166.
- Report of the Household Income and Expenditure Survey* (2005), Bangladesh Bureau of Statistics (BBS), Planning Division, Ministry of Planning, Government of the People's Republic of Bangladesh, May 2007.
- Report of the Household Income and Expenditure Survey* (2010), Bangladesh Bureau of Statistics (BBS), Planning Division, Ministry of Planning, Government of the People's Republic of Bangladesh, February 2010.
- Sahu, S. K., & Smith, T. M. F. (2006). A Bayesian method of sample size determination with practical applications. *Royal Statistical Society*, 169(Part 2), 235-253.

Estimation of Gumbel Parameters under Ranked Set Sampling

Omar M. Yousef
Al Balqa' Applied University
Zarqa, Jordan

S. A. Al-Subh
Mutah University
Karak, Jordan

Consider the MLEs (maximum likelihood estimators) of the parameters of the Gumbel distribution using SRS (simple random sample) and RSS (ranked set sample) and the MOMEs (method of moment estimators) and REGs (regression estimators) based on SRS. A comparison between these estimators using bias and MSE (mean square error) was performed using simulation. It appears that the MLE based on RSS can be a robust competitor to the MLE based on SRS.

Keywords: Ranked set sampling; simple random sampling, parameters, Gumbel distribution, maximum likelihood estimator, bias, mean square error, regression estimator, method of moment estimator.

Introduction

There are many areas of application of the Gumbel distribution including environmental sciences, system reliability, and hydrology. In hydrology, for example, the Gumbel distribution may be used to represent the distribution of the minimum level of a river in a particular year based on minimum values for the past few years. It is useful for predicting the occurrence of extreme earthquakes, floods, and other natural disasters. The potential applicability of the Gumbel distribution to represent the distribution of minima relates to extreme value theory, which indicates that it is likely to be useful if the distribution of the underlying sample data is of the normal or exponential type.

The problem of estimation of the unknown parameters of the Gumbel distribution is considered by many authors under simple random sampling. Maciunas et al. (1979) considered the estimates of the parameters of the Gumbel distribution by the methods of probability weighted moments, moments, and

Omar M. Yousef is a lecturer in the Basic Sciences Department. Email him at abuyazan_jo@yahoo.com. Sameer A. Al-Subh is an assistant professor in the Department of Mathematics and Statistics. Email him at salsubh@yahoo.com.

maximum likelihood. They used both independent and serially correlated Gumbel numbers to derive the results from Monte Carlo experiments. They found the method of probability weighted moments estimator is more efficient than the estimators. Leese (1973), derived the MLE (maximum likelihood estimator) of Gumbel distribution parameters in case of censored samples and he gave expressions for their large-sample standard errors. Fiorentino and Gabriele (1984), given some modifications of the MLE the Gumbel distribution parameters to reduce the bias of the estimators. Phien (1987) estimated the parameters of the Gumbel distribution by moments, MLE, maximum entropy and probability weighted moments. He derived the asymptotic variance-covariance matrix of the MLEs and used simulation to compare between the various estimators. He found that the MLE is best in terms of the root MSE (mean square error). Corsini et al. (1995), discussed the MLE and Cramer-Rao (CR) bounds for the location and scale parameters of the Gumbel distribution. Mousa et al. (2002), found the Bayesian estimation for the two parameters of the Gumbel distribution based on record values.

RSS as introduced by McIntyre (1952) is an ingenious sampling technique for selecting a sample which is more informative than a SRS to estimate the population mean. He used of RSS technique to estimate the mean pasture and forage yields. RSS technique is very useful when visual ranking of population units is less expensive than their actual quantifications. Therefore, selecting a sample based on RSS can reduce the cost and increase the efficiency of estimation.

The basic idea behind selecting a sample under RSS can be described as follows: Select m random samples each of size m , using a visual inspection or any cheap method to rank the units within each sample with respect to the variable of interest. Then select, for actual measurement, the i^{th} smallest unit from the i^{th} sample, $i = 1, \dots, m$. In this way, a total of m measured units are obtained, one from each sample. The procedure could be repeated r times until a sample of $n = mr$ measurements are obtained. These mr measurements form RSS. Takahasi and Wakimoto (1968) gave the theoretical background for RSS. They showed that the mean of an RSS is an unbiased estimator of the population mean with variance smaller than that of the mean of a SRS. Dell and Clutter (1972) showed that the RSS mean remains unbiased and more efficient than the SRS mean for estimating the population even if ranking is not perfect. A comprehensive survey about developments in RSS can be found in Chen (2000) and Muttalak and Al-Saleh (2000).

Because there are many attractive applications of Gumbel distribution, it is of interest to conduct a statistical inference for the Gumbel distribution. The statistical inference includes the study of some properties of Gumbel distribution,

emphasizing on estimation of Gumbel parameters. The estimation of the location and scale parameters, denoted as α and β respectively, of the Gumbel distribution under SRS and RSS is studied. The Gumbel parameters were estimated by using several methods of estimation in both cases of SRS and RSS such as maximum likelihood, moments and regression. Furthermore, the performance of these estimators is investigated and compared through simulation. Bias, mean square error (MSE) and efficiency of these estimators were used for comparison.

Parameter Estimation Using SRS

The cdf and pdf of the random variable X which has a Gumbel distribution with parameters α and β are given respectively by

$$F(x; \alpha, \beta) = \exp\left(-\exp\left(-\frac{x-\alpha}{\beta}\right)\right), \quad (1)$$

$$f(x; \alpha, \beta) = \frac{1}{\beta} \exp\left(-\frac{x-\alpha}{\beta} - \exp\left(-\frac{x-\alpha}{\beta}\right)\right), \quad (2)$$

where α is the location parameter and β is the scale parameter, $\beta > 0$, x and $\alpha \in (-\infty, \infty)$.

Let X_1, X_2, \dots, X_n be a random sample from X . The MLEs, MOMEs (method of moment estimators) and REGs (regression estimators) will be examined in case of both parameters are unknown based on X_1, X_2, \dots, X_n .

MLEs

Let X_1, X_2, \dots, X_n be a random sample from (2). The log-likelihood function is given by

$$l(\alpha, \beta) = -n \log(\beta) - \sum_{i=1}^n \left(\frac{x_i - \alpha}{\beta} \right) - \sum_{i=1}^n \exp\left(-\frac{x_i - \alpha}{\beta}\right). \quad (3)$$

After taking the derivatives with respect to α and β equating to 0, the MLEs are obtained as

$$\hat{\beta}_{MLE,S} = \bar{x} - \sum_{i=1}^n x_i w_i \text{ and } \hat{\alpha}_{MLE,S} = -\hat{\beta}_{MLE,S} \log(\bar{z}). \quad (4)$$

where $z_i = \exp\left(-\frac{x_i}{\hat{\beta}_{MLE,S}}\right)$, $\bar{z} = \frac{1}{n} \sum_{i=1}^n z_i$ and $w_i = \frac{z_i}{n\bar{z}}$.

MOMEs

The mean and variance for Gumbel distribution are given by

$$\mu = \alpha + \gamma\beta \text{ and } \sigma^2 = \frac{\pi^2}{6} \beta^2. \quad (5)$$

The moment estimators of the two parameters are

$$\hat{\beta}_{MOME,S} = \frac{\sqrt{6}}{\pi} s \text{ and } \hat{\alpha}_{MOME,S} = \bar{x} - \gamma \hat{\beta}_{MOME,S} \quad (6)$$

where s , \bar{x} are the sample standard deviation and mean, respectively, and $\gamma = 0.57721566$ is Euler's constant.

REGs

$$\begin{aligned} \text{Let } y = F(x; \alpha, \beta) &= \exp\left(-\exp\left(-\frac{x-\alpha}{\beta}\right)\right) \Rightarrow \ln y = -\exp\left(-\frac{x-\alpha}{\beta}\right) \\ \Rightarrow -\ln y &= \exp\left(-\frac{x-\alpha}{\beta}\right) \Rightarrow t = \ln(-\ln y) = -\left(\frac{x-\alpha}{\beta}\right) \Rightarrow t = a x + b \\ \text{where } a &= \frac{-1}{\beta} \text{ and } b = \frac{\alpha}{\beta}. \end{aligned}$$

ESTIMATION OF GUMBEL PARAMETERS

The regression estimators of the two parameters are

$$\hat{\beta}_{REG,S} = \frac{-1}{\hat{a}} \text{ and } \hat{\alpha}_{REG,S} = \hat{\beta}_{REG,S} (t - \hat{a}\bar{x}) \quad (7)$$

$$\text{where } \hat{a} = \frac{\sum_{i=1}^n (x_i - \bar{x})(t_i - \bar{t})}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Parameter Estimation Under RSS

MLEs

Let $X_{(i:m)j}$, $i = 1, \dots, m$ and $j = 1, \dots, r$ denote the i^{th} order statistics from the i^{th} set of size m of the j^{th} cycle be the RSS data for X with sample size $n = mr$.

Using (1) and (2), the pdf of $X_{(i:m)j}$ is given by (Arnold et al., 1992)

$$f_{i:m}(X_{(i:m)j}) = c \left(F(X_{(i:m)j}) \right)^{i-1} \left(1 - F(X_{(i:m)j}) \right)^{m-i} f(X_{(i:m)j}), \text{ where } c = \frac{1}{B(i, m-i+1)},$$

$$f(X_{(i:m)j}) = \frac{1}{\beta} \exp \left(- \left(\frac{X_{(i:m)j} - \alpha}{\beta} \right) - \exp \left(- \left(\frac{X_{(i:m)j} - \alpha}{\beta} \right) \right) \right) \text{ and}$$

$$F(X_{(i:m)j}) = \exp \left(- \exp \left(- \left(\frac{X_{(i:m)j} - \alpha}{\beta} \right) \right) \right). \text{ Then the likelihood function is given}$$

by

$$\begin{aligned} l(\alpha, \beta) &= \prod_{i=1}^r \prod_{j=1}^m f_{i:m}(X_{(i:m)j}) \\ &= \prod_{i=1}^r \prod_{j=1}^m \left[c \left(F(X_{(i:m)j}) \right)^{i-1} \left(1 - F(X_{(i:m)j}) \right)^{m-i} f(X_{(i:m)j}) \right] \\ &= c^{mr} \prod_{i=1}^r \prod_{j=1}^m \left[\left(F(X_{(i:m)j}) \right)^{i-1} \left(1 - F(X_{(i:m)j}) \right)^{m-i} f(X_{(i:m)j}) \right]. \end{aligned}$$

The log-likelihood function is given by

$$L(\alpha, \beta) = mr \log c + \sum_{j=1}^r \sum_{i=1}^m (i-1) \log F(X_{(i:m)j}) + \sum_{j=1}^r \sum_{i=1}^m (m-i) \log(1 - F(X_{(i:m)j})) + \sum_{j=1}^r \sum_{i=1}^m \log f(X_{(i:m)j}). \quad (8)$$

Taking the derivatives of (8) with respect to α and β respectively, and equating the resulting quantities to zero. Because there is no explicit solution for (8), the equations need to be solved numerically to find $\hat{\alpha}_{MLE,R}$ and $\hat{\beta}_{MLE,R}$.

Ad-hoc Estimators

These are the same as the estimators in (6) and (7) with SRS replaced by RSS

Estimator Comparison

A comparison between all above estimators for both parameters of the Gumbel distribution was carried out under SRS and RSS using simulation. The package R has been used to conduct the simulation. The following values of the parameters and sample sizes have been considered: $\alpha = 0.5, \beta = 1$; $\alpha = 1, \beta = 0.5$; $\alpha = 1, \beta = 1$; $\alpha = 1, \beta = 2$; $\alpha = 2, \beta = 1$, $n = 12$ and $n = 24$.

For each n , a set $(m;r)$ is decided such that $n = mr$. The bias and the MSE are computed for all the estimators under consideration. The efficiency between all estimators with respect to the MLE based on SRS are calculated where the efficiency of the estimator is defined as

$$eff(\hat{\theta}_2, \hat{\theta}_1) = \frac{MSE(\hat{\theta}_1)}{MSE(\hat{\theta}_2)}$$

where

$$MSE(\hat{\theta}) = \frac{1}{10,000} \sum_{t=1}^{10,000} (\hat{\theta}_{2t} - \theta_2)^2.$$

If $eff(\hat{\theta}_2, \hat{\theta}_1) > 1$ then $\hat{\theta}_2$ is better than $\hat{\theta}_1$.

The bias of the estimators is reported in Tables 1 and 3 and the efficiencies of the estimators is reported in Tables 2 and 4.

ESTIMATION OF GUMBEL PARAMETERS

Table 1. The bias and MSE of estimators of α

(α, b)	n	$n=mr$	Bias					MSE				
			$\hat{\alpha}_{mle,S}$	$\hat{\alpha}_{moe,S}$	$\hat{\alpha}_{reg,S}$	$\hat{\alpha}_{mle,R}$	$\hat{\alpha}_{moe,R}$	$\hat{\alpha}_{mle,S}$	$\hat{\alpha}_{moe,S}$	$\hat{\alpha}_{reg,S}$	$\hat{\alpha}_{mle,R}$	$\hat{\alpha}_{moe,R}$
(1,1)	12	$m=2, r=6$			0.121	-1.156	-1.141	1.647	1.499	0.058	1.522	1.486
		$m=3, r=4$	-1.183	-1.130	-1.153		-1.146				1.480	1.469
		$m=4, r=3$			-1.140		-1.146				1.427	1.444
	24	$m=2, r=12$			-1.187	1.601	-1.149	1.601	1.424	0.033	1.515	1.415
		$m=3, r=8$	-1.213	0.103	-1.179		-1.152				1.473	1.405
		$m=4, r=6$			-1.167		-1.152				1.430	1.393
(1,2)	12	$m=2, r=6$			2.268	5.841	-2.285	5.841	5.907	2.205	4.212	5.943
		$m=3, r=4$	-1.728	-1.471	2.174		-2.283				4.211	5.799
		$m=4, r=3$			2.092		-2.292				4.211	5.793
	24	$m=2, r=12$			-2.047	4.211	-2.298	4.211	5.666	2.296	5.810	5.636
		$m=3, r=8$	2.321	-1.508	2.135		-2.306				4.211	5.628
		$m=4, r=6$			2.174		-2.302				4.211	5.567
(2,1)	12	$m=2, r=6$			-1.167	1.666	-1.144	1.666	1.519	3.793	1.549	1.486
		$m=3, r=4$	-1.728	1.901	-1.160		-1.141				1.499	1.446
		$m=4, r=3$			-1.153		-1.151				1.456	1.457
	24	$m=2, r=12$			-1.203	1.640	-1.140	1.640	1.432	3.793	1.549	1.389
		$m=3, r=8$	-1.229	1.921	-1.182		-1.148				1.480	1.395
		$m=4, r=6$			-1.176		-1.152				1.453	1.394
(0.5,1)	12	$m=2, r=6$			-1.139	1.558	-1.144	1.558	1.516	0.362	1.485	1.490
		$m=3, r=4$	-1.143	-0.576	-1.131		-1.141				1.427	1.457
		$m=4, r=3$			-1.129		-1.148				1.410	1.448
	24	$m=2, r=12$			-1.168	1.557	-1.140	1.557	1.432	0.389	1.469	1.392
		$m=3, r=8$	-1.185	-0.612	-1.159		-1.142				1.436	1.381
		$m=4, r=6$			-1.137		-1.149				1.371	1.388
(1,0.5)	12	$m=2, r=6$			-0.589	0.419	-0.571	0.419	0.376	0.205	0.393	0.371
		$m=3, r=4$	-0.598	0.375	-0.575		-0.569				0.368	0.361
		$m=4, r=3$			-0.577		-0.572				0.364	0.360
	24	$m=2, r=12$			-0.602	0.419	-0.576	0.419	0.353	0.157	0.387	0.356
		$m=3, r=8$	-0.619	0.344	-0.597		-0.573				0.378	0.347
		$m=4, r=6$			-0.585		-0.575				0.359	0.347

Table 2. The efficiency of estimators of α

(α, b)	n	$n=mr$	$\hat{\alpha}_{mle,S}$	$\hat{\alpha}_{moe,S}$	$\hat{\alpha}_{reg,S}$	$\hat{\alpha}_{mle,R}$	$\hat{\alpha}_{moe,R}$
(1,1)	12	$m=2, r=6$	1	1.099	28.397	1.082	1.108
		$m=3, r=4$				1.113	1.121
		$m=4, r=3$				1.154	1.141
	24	$m=2, r=12$	1	1.124	48.515	1.057	1.131
		$m=3, r=8$				1.087	1.140
		$m=4, r=6$				1.120	1.149
(1,2)	12	$m=2, r=6$	1	0.989	2.649	1.387	0.983
		$m=3, r=4$				1.387	1.007
		$m=4, r=3$				1.387	1.008
	24	$m=2, r=12$	1	0.743	1.834	0.725	0.747
		$m=3, r=8$				1.000	0.748
		$m=4, r=6$				1.000	0.756
(2,1)	12	$m=2, r=6$	1	1.097	0.439	1.076	1.121
		$m=3, r=4$				1.111	1.152
		$m=4, r=3$				1.144	1.143
	24	$m=2, r=12$	1	1.145	0.432	1.059	1.181
		$m=3, r=8$				1.108	1.176
		$m=4, r=6$				1.129	1.176
(0.5,1)	12	$m=2, r=6$	1	1.028	4.304	1.049	1.046
		$m=3, r=4$				1.092	1.069
		$m=4, r=3$				1.105	1.076
	24	$m=2, r=12$	1	1.087	4.003	1.060	1.119
		$m=3, r=8$				1.084	1.127
		$m=4, r=6$				1.136	1.122
(1,0.5)	12	$m=2, r=6$	1	1.114	2.044	1.066	1.129
		$m=3, r=4$				1.139	1.161
		$m=4, r=3$				1.151	1.164
	24	$m=2, r=12$	1	1.187	2.669	1.083	1.177
		$m=3, r=8$				1.108	1.207
		$m=4, r=6$				1.167	1.207

ESTIMATION OF GUMBEL PARAMETERS

Table 3. The bias and MSE of estimators of β

(α, b)	n	$n=mr$	Bias					MSE				
			$\hat{\alpha}_{mle,S}$	$\hat{\alpha}_{moe,S}$	$\hat{\alpha}_{reg,S}$	$\hat{\alpha}_{mle,R}$	$\hat{\alpha}_{moe,R}$	$\hat{\alpha}_{mle,S}$	$\hat{\alpha}_{moe,S}$	$\hat{\alpha}_{reg,S}$	$\hat{\alpha}_{mle,R}$	$\hat{\alpha}_{moe,R}$
(1,1)	12	$m=2, r=6$			0.889	0.296	-0.023	0.327	0.077	0.963	0.337	0.079
		$m=3, r=4$	0.273	-0.042		0.303	-0.013				0.301	0.074
		$m=4, r=3$				0.308	-0.011				0.302	0.069
	24	$m=2, r=12$			0.919	0.415	-0.012	0.359	0.042	0.949	0.353	0.043
		$m=3, r=8$	0.412	-0.021		0.425	-0.006				0.348	0.039
		$m=4, r=6$				0.416	0.006				0.322	0.036
(1,2)	12	$m=2, r=6$			4.368	1.737	0.959	5.293	1.154	19.696	4.759	1.234
		$m=3, r=4$	1.772	0.918		1.734	0.959				4.508	1.196
		$m=4, r=3$				1.714	0.986				4.128	1.253
	24	$m=2, r=12$			4.445	2.018	0.974	5.461	1.081	20.192	5.127	1.106
		$m=3, r=8$	2.328	0.955		1.928	0.981				4.656	1.119
		$m=4, r=6$				1.734	0.984				4.509	1.113
(2,1)	12	$m=2, r=6$			0.799	-0.707	-1.022	0.773	1.155	0.877	0.728	1.122
		$m=3, r=4$	-0.740	-1.039		-0.693	-1.020				0.695	1.111
		$m=4, r=3$				-0.688	-1.010				0.674	1.091
	24	$m=2, r=12$			0.850	-0.584	-1.016	0.533	1.082	0.862	0.514	1.072
		$m=3, r=8$	-0.599	-1.019		-0.589	-1.011				0.490	1.061
		$m=4, r=6$				-0.587	-1.007				0.483	1.051
(0.5,1)	12	$m=2, r=6$			0.979	0.803	0.478	0.923	0.286	1.102	0.896	0.304
		$m=3, r=4$	0.790	0.455		0.811	0.485				0.891	0.311
		$m=4, r=3$				0.808	0.488				0.875	0.306
	24	$m=2, r=12$			1.001	0.927	0.481	1.045	0.274	1.089	1.040	0.272
		$m=3, r=8$	0.913	0.482		0.927	0.489				1.035	0.276
		$m=4, r=6$				0.908	0.492				0.974	0.277
(1,0.5)	12	$m=2, r=6$			-0.257	-0.356	-0.512	0.193	0.291	0.119	0.181	0.282
		$m=3, r=4$	-0.369	-0.522		-0.358	-0.512				0.179	0.280
		$m=4, r=3$				-0.347	-0.506				0.169	0.272
	24	$m=2, r=12$			-0.265	-0.291	-0.504	0.129	0.272	0.105	0.126	0.266
		$m=3, r=8$	-0.292	-0.512		-0.289	-0.506				0.123	0.265
		$m=4, r=6$				-0.295	-0.505				0.122	0.264

Table 4. The efficiency of estimators of β

(α, b)	n	$n=mr$	$\hat{\alpha}_{mle,S}$	$\hat{\alpha}_{moe,S}$	$\hat{\alpha}_{reg,S}$	$\hat{\alpha}_{mle,R}$	$\hat{\alpha}_{moe,R}$
(1,1)	12	$m=2, r=6$	1	4.247	0.340	0.973	4.139
		$m=3, r=4$				1.086	4.419
		$m=4, r=3$				1.083	4.739
	24	$m=2, r=12$	1	8.548	0.378	1.017	8.349
		$m=3, r=8$				1.032	9.205
		$m=4, r=6$				1.115	9.972
(1,2)	12	$m=2, r=6$	1	4.587	0.269	1.112	4.289
		$m=3, r=4$				1.174	4.426
		$m=4, r=3$				1.282	4.224
	24	$m=2, r=12$	1	5.052	0.270	1.065	4.938
		$m=3, r=8$				1.173	4.880
		$m=4, r=6$				1.211	4.907
(2,1)	12	$m=2, r=6$	1	0.669	0.881	1.062	0.689
		$m=3, r=4$				1.112	0.696
		$m=4, r=3$				1.147	0.709
	24	$m=2, r=12$	1	0.493	0.618	1.037	0.494
		$m=3, r=8$				1.088	0.500
		$m=4, r=6$				1.104	0.504
(0.5,1)	12	$m=2, r=6$	1	3.227	0.838	1.030	3.036
		$m=3, r=4$				1.036	2.968
		$m=4, r=3$				1.055	3.016
	24	$m=2, r=12$	1	3.814	0.960	1.005	3.842
		$m=3, r=8$				1.010	3.786
		$m=4, r=6$				1.073	3.773
(1,0.5)	12	$m=2, r=6$	1	0.663	1.622	1.066	0.684
		$m=3, r=4$				1.078	0.689
		$m=4, r=3$				1.142	0.710
	24	$m=2, r=12$	1	0.474	1.229	1.024	0.485
		$m=3, r=8$				1.049	0.487
		$m=4, r=6$				1.057	0.489

From Tables 1 to 4, the following conclusions are put forth

- In general, the bias is large for all estimators. Therefore, all the estimators are considered as biased estimators for α .
- From Table 1, it can be noticed that the REG under SRS has the smallest bias as compared to the other estimators considered in most cases. In general, for all estimators of α under RSS, the bias is less than the case under SRS.
- For fixed α , the MSE of $\hat{\alpha}$ decreases as the sample size increases.
- It is noticed that from Table 2 that MLE under RSS is the most efficient than the MLE based on SRS.

ESTIMATION OF GUMBEL PARAMETERS

- v) The efficiency of the other estimators (MOMEs and REGs based on SRS and RSS) are not consistent, sometimes less one and other times greater than 1.

Similar remarks can be noticed for the case of β .

Conclusion

Based on this study, it may be concluded that all estimators are biased. Because the MLEs under RSS are more efficient than the MLE under SRS, RSS is recommended in case ordering can be done visually or by an inexpensive method. The other estimators are not recommended because they are not consistent.

References

- Arnold, B. C., Balakrishnan, N., & Nagaraja, H. N. (1992). *A First Course in Order Statistics*. New York: John Wiley and Sons.
- Chen, Z. (2000). On ranked-set sample quantiles and their applications. *Journal of Statistical Planning and Inference*, 83, 125-135.
- Corsini, G., Gini, F., Greco, M. V., & Verrazzani, L. (1995). Cramer-Rao bounds and estimation of the parameters of the Gumbel distribution. *Aerospace and Electronic Systems*, 31(3), 1202-1204.
- Dell, D. R., & Clutter, J. L. (1972). Ranked set sampling theory with order statistics background. *Biometrics*, 28, 545-555.
- Fiorentino, M., & Gabriele, S. (1984). A correction for the bias of maximum-likelihood estimators of Gumbel parameters. *Journal of Hydrology*, 73(1-2), 39-49.
- Gumbel, E. J. (1958). *Statistics of Extremes*. New York: Columbia University Press.
- Leese, M. N. (1973). Use of censored data in the estimation of Gumbel distribution parameters for annual maximum flood series. *Water Resources Research*, 9(6), 1534-1542. doi: [10.1029/WR009i006p01534](https://doi.org/10.1029/WR009i006p01534)
- Maciunas, J., Matalas, N. C., & Wallis, J. R. (1979). Probability weighted moments compared with some traditional techniques in estimating Gumbel Parameters and quantiles. *Water Resources Research*, 15(5), 1055-1064. doi: [10.1029/WR015i005p01049](https://doi.org/10.1029/WR015i005p01049)

- McIntyre, G. A. (1952). A method of unbiased selective sampling using ranked sets. *Australian Journal of Agricultural Research*, 3, 385-390.
- Mousa, M. A. M., Jaheen, Z. F., & Ahmad, A. A. (2002). Bayesian Estimation, Prediction and Characterization for the Gumbel Model Based on Records. *A Journal of Theoretical and Applied Statistics*, 36(1), 65-74.
- Muttlak, H. A., & Al-Saleh, M. F. 2000. Recent developments in ranked set sampling, *Pakistan Journal of Statistics*, 16, 269-290.
- Phien, H. N. (1987). A review of methods of parameter estimation for the extreme value type-1 distribution. *Journal of Hydrology*, 90(3-4), 251-268
- Takahasi, K., & Wakitmoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Annals of the Institute of Statistical Mathematics*, 20, 1-31.

The Information Criterion

M. Ghahramani

Payam Noor University
Tehran, Iran

The Akaike information criterion, AIC, is widely used for model selection. Using the AIC as the estimator of asymptotic unbiased for the second term Kullback-Leibler risk considers the divergence between the true model and offered models. However, it is an inconsistent estimator. A proposed approach the problem is the use of AIC, a consistently offered information criterion. Model selection of classic and linear models are considered by a Monte Carlo simulation.

Keywords: Consistency, AIC, information criterion, Kullback-Leibler risk, model selection

Introduction

Statistical modeling is used for investigating a random phenomenon that isn't completely predictable. One of the criteria frequently used in model selection is the Kullback-Leibler (KL) information criterion (Kullback and Leibler, 1951). This information criterion was introduced as one risk in model selection. Akaike (1973) introduced information criterion, AIC, as an estimator of asymptotic unbiased for the second term KL risk and to form a penalty likelihood function. Akaike stated modeling isn't only finding a model which describes the behavior of the observed data, but its main aim is predicated as a possible good, and the future of the process is under investigation. Hall (1987) used the Kullback-Leibler risk considered bias and variance in the approximate density function. Bozdogan (2000), with the error distinction in the model selection, considered two errors from bias and variance in the estimation of model selection. Choi and Kiffer (2006), and Cawley and Talbot (2010) have considered the over fitting in model selection, and they showed over fitting results from the bias when modeling phenomena have been considered. Over the years, corrections have been made on penalty term, and criteria such as AIC (Akaike, 1973), TIC (Takeuchi, 1976), and KIC (Cavanaugh, 1994) have been

M. Ghahramani is in the Department of Mathematical Statistics. Email him at estimatormgh@gmail.com.

introduced. In section 2, we state the Kullbake-Liebler risk, and the necessity of definitions. In section 3, a consistent information criterion is proposed instead of the AIC. In section 4, we present the results of our simulation studies.

Kullbake-Leibler Risk

Let $X = (X_1, X_2, \dots, X_n)$ be a (i.i.d) random sample from true model and unknown, $h(\cdot)$, and the family $F_{\theta_k} = \{f(\cdot; \theta_k) = f_{\theta_k}; \theta_k \in \Theta \subseteq R^k\}$ from offered models has been considered for approximate true model.

Definition 1

The family F_{θ_k} is well specified if there is a $\theta_0 \in \Theta$ such that $h(\cdot) = f(\cdot; \theta_0)$; otherwise it is misspecified.

Definition 2

The KL risk defines for generate model and unknown $h(\cdot)$, and offered model f_{θ_k} as

$$KL(h, f_{\theta_k}) = E_h \left[\log \left(\frac{h(\cdot)}{f(\cdot; \theta_k)} \right) \right] = E_h[\log h(\cdot)] - E_h[\log f(\cdot; \theta_k)] \quad (1)$$

where the expectation is taken with respect to the unknown model $h(\cdot)$. The first term in the right hand side of (1) is called irrelevant part, because it doesn't depend on θ_k , and the second term is called relevant part. Based on the properties of the KL risk, the smaller value showed the closeness of the offered model to the unknown and true model. Therefore the problem is reduced to obtain a good estimate of the expected log-likelihood. Since the expectation is with respect to the model with unknown parameters, one estimator is

$$E_h \{ \log f(\cdot; \hat{\theta}_n) \} = \frac{1}{n} \sum_{i=1}^n \log f(x_i; \hat{\theta}_n).$$

Thus, $\hat{\theta}_n$ is the maximum likelihood estimator of θ_k and $f(\cdot; \hat{\theta}_n)$ is the maximum likelihood function. The bias of maximum log-likelihood is as

PAIRWISE COMPARISON IN REPEATED MEASURES

$$\text{Bias estimator} = E_h \{ \log f(.; \hat{\theta}_n) - nE_h \{ \log f(Z ; \hat{\theta}_n) \}$$

where Z is a random variable (i.i.d) with X_i 's. The general form of the information criterion that has been shown by IC, as

$$IC = -2\sum_{i=1}^n \log f(X_i ; \hat{\theta}_n) + 2\{\text{bias estimator}\} = -2l_f(\hat{\theta}_n) + 2\{\text{bias estimator}\}.$$

Akaike, when offered family is well specified, size of bias is estimated with dimensional parameter $\hat{\theta}_n$, means k , and the Akaike information criterion is stated as

$$AIC = -2l_f(\hat{\theta}_n) + 2k.$$

With attention to form the AIC by increasing the number of parameters in the offered model the penalty term, $2k$ will be increased and the term $-2\sum_{i=1}^n \log f(X_i ; \hat{\theta}_n)$ will be decrease. Penalty term is constant to chance of size sample in the information criterion AIC, and by increasing the size sample, AIC cannot distinguish the true model with the probability one. Therefore this problem is the same concept of inconsistency for an information criterion. Following the inconsistency of information criterion AIC, based on the definition similar to the definition of AIC, a consistent of information criterion, which called A'IC has presented. Akaike information criterion, by Akaike for model selection is introduced, but this useful criterion is inconsistent ([Akaike, 1973](#)).

The information criterion is obtained as follows. The basis of the log-likelihood function is

$$b = E_h \{ \log f(., \hat{\theta}_n) - nE_h \log f(Z ; \hat{\theta}_n) \}$$

where in the second term of the right hand side the inner expectation is calculated with respect to $h(z)$ and the outer expectation is calculated with respect to $h(x)$. By evaluating the bias it is decomposed as follows:

$$\begin{aligned} b &= E_h \{ \log f(., \hat{\theta}_n) - \log f(., \theta_0) \} + E_h \{ \log f(., \theta_0) - nE_h \{ \log f(Z ; \theta_0) \} \} \\ &\quad + nE_h \{ E_h \{ \log f(Z ; \theta_0) - E_h \{ \log f(Z ; \hat{\theta}_n) \} \} \} = b_1 + b_2 + b_3. \end{aligned}$$

The three expectations are calculated separately b_1 , b_2 , and b_3 .

a) For calculation of b_1 by writing $l_f(\theta_0) = \log f(., \theta_0)$ and by applying a Taylor series expansion around the maximum likelihood estimator $\hat{\theta}_n$, results in

$$l_f(\theta_0) = l_f(\hat{\theta}_n) + (\theta_0 - \hat{\theta}_n)^T \frac{\partial l_f(\theta)}{\partial \theta} \Big|_{\theta=\hat{\theta}_n} + \frac{1}{2} (\theta_0 - \hat{\theta}_n)^T \frac{\partial^2 l_f(\theta)}{\partial \theta \partial \theta^T} \Big|_{\theta=\hat{\theta}_n} (\theta_0 - \hat{\theta}_n) + o_p(1), \quad (1)$$

$O_p(1)$ is an expression of quantity that in the probability tends to zero.

With attention, the $\frac{\partial l_f(\theta)}{\partial \theta} \Big|_{\theta=\hat{\theta}_n} = 0$ and $\frac{1}{n} \frac{\partial^2 l_f(\theta)}{\partial \theta \partial \theta^T} \Big|_{\theta=\hat{\theta}_n}$ is converged to $J(\theta_0)$. (Akaike, 1973). Thus,

$$J(\theta_0) = -E_h \left[\frac{\partial^2 l_f(\theta)}{\partial \theta \partial \theta^T} \right] \Big|_{\theta=\theta_0}$$

Thus, the relation above can be approximated, as

$$l_f(\hat{\theta}_n) - l_f(\theta_0) \approx \frac{n}{2} (\theta_0 - \hat{\theta}_n)^T J(\theta_0) (\theta_0 - \hat{\theta}_n) + o_p(1),$$

This based on the b_1 can be written as

$$b_1 = E_h \{ l_f(\hat{\theta}_n) - l_f(\theta_0) \} \approx E_h \left\{ \frac{n}{2} (\theta_0 - \hat{\theta}_n)^T J(\theta_0) (\theta_0 - \hat{\theta}_n) \right\} \quad (2)$$

b) The b_2 doesn't contain an estimator and it can easily be written as

$$b_2 = E_h \{ g f(., \theta_0) - n E_h \{ \log f(Z; \theta_0) \} \} = 0 \quad (3)$$

c) For calculation of value the b_3 , first, the phrase $E_h \{ \log f(Z; \theta_0) \}$ be defined equally of $\eta(\hat{\theta}_n)$. By using from Taylor expectation $\eta(\hat{\theta}_n)$ around θ_0 ,

$$\eta(\hat{\theta}_n) = \eta(\theta_0) + (\hat{\theta}_n - \theta_0)^T \frac{\partial \eta(\theta)}{\partial \theta} \Big|_{\theta=\theta_0} + \frac{1}{2} (\hat{\theta}_n - \theta_0)^T \frac{\partial^2 \eta(\theta)}{\partial \theta \partial \theta^T} \Big|_{\theta=\theta_0} (\hat{\theta}_n - \theta_0) + o_p(1)$$

PAIRWISE COMPARISON IN REPEATED MEASURES

with attention to the $\frac{\partial \eta(\theta)}{\partial \eta}|_{\theta=\theta_0} = 0$. Thus when n tends to infinity, the relation above can be approximated as

$$\eta(\hat{\theta}_n) \approx \eta(\theta_0) + \frac{1}{2}(\hat{\theta}_n - \theta_0)^T J(\theta_0)(\hat{\theta}_n - \theta_0) + o_p(1).$$

Thus the b_3 can be written as

$$\begin{aligned} b_3 &= nE_h \{E_h \{\log f(Z; \theta_0)\} - E_h \{\log f(Z; \hat{\theta}_n)\}\} \\ &\approx \frac{n}{2} E_h \{(\theta_0 - \hat{\theta}_n)^T J(\theta_0)(\hat{\theta}_n - \theta_0)\} \end{aligned} \quad (4)$$

If the family of F_{θ_k} is well specified, with attention to quadratic forms in relations (2) and (4), that converge to centrally distributed chi-square with k degrees of freedom, then b_1 and b_3 can be written as

$$b_1 = b_3 = \frac{n}{2} k \quad (5)$$

By combining of b_1 and b_3 , in relation (5) and b_2 , in relation (3), bias the b is $b = b_1 + b_2 + b_3 = nk$.

By replacing the value of b in the general form of the information criterion, the offered information criterion called, A'IC is obtained as

$$A'IC = -2\sum_{i=1}^n \log f(X_i; \hat{\theta}_n) + 2nk \quad (6)$$

In the offered information criterion A'IC, penalty term $2nk$ changes will change with sample size. So, if sample size will be very large, information criterion A'IC, with the probability of one, find the true model data. In other words, information criterion A'IC is the only consistent information criterion that has been obtained based on the Kullback-Leibler risk. To show consistency of information criterion A'IC, let the maximum likelihood function estimator for the offered model $(f(\cdot; \theta_k) = f(\theta_k))$ and optimal model $(f(\cdot; \theta_{k_0}) = f(\theta_{k_0}))$ with respectively

$l_f(\hat{\theta}_{k(n)})$ and $l_f(\hat{\theta}_{k_0(n)})$. With regard to relation (6) information criterion A'IC, for the model $f(\theta_k)$ and $f(\theta_{k_0})$, we have

$$A'IC(f(\theta_k)) = -2l_f(\hat{\theta}_{k(n)}) + 2nk,$$

$$A'IC(f(\theta_{k_0})) = -2l_f(\hat{\theta}_{k_0(n)}) + 2nk_0$$

If there is $k > k_0$, consistency for information criterion A'IC is given by

$$\begin{aligned} &P(A'ICf(\theta_k) - A'ICf(\theta_{k_0}) > 0) \\ &= P(-2l_f(\hat{\theta}_{k(n)}) + 2nk - (-2l_f(\hat{\theta}_{k_0(n)}) + 2nk_0) > 0) \\ &= P(2l_f(\hat{\theta}_{k(n)}) - 2l_f(\hat{\theta}_{k_0(n)}) < 2nk - 2nk_0) \\ &= P(U_n < 2n(k - k_0)) = F(2n(k - k_0))^p \rightarrow F(\infty) = 1 \quad (7) \end{aligned}$$

In relation (7), U_n is $2l_f(\hat{\theta}_{k(n)}) - 2l_f(\hat{\theta}_{k_0(n)})$ and the distribution function of chi-square has been shown by F. Therefore it tends in of the probability to one. Thus A'IC is a consistent information criterion. (For further study about the consistency of an information criterion, see [Hu and Shao 2008](#)).

Simulation

A simulation was conducted for usage and comparison of the offered information criterion, A'IC, with the information criterion AIC, by using Monte-Carlo simulation, for linear regression and classic models. This simulation of linear regression model is supposed that well specified family $F_{\theta_k} = \{f(\cdot; \theta_k) = f_{\theta_k}; \theta_k \in \Theta \subseteq R^k\}$, and misspecified family $G_{\beta_d} = \{g(\cdot; \beta_d) = g_{\beta_d}; \beta_d \in B \subseteq R^d\}$ are given for estimating the true model. Let $f: y_i = 0.3 + 0.5x_{i1} + x_{i2} + 0.7x_{i3} + \varepsilon_{i1} i = 1, \dots, n$ as the true model so that ε_{i1} , has been generated as random from distribution $N(0, 2)$. Models $f_1: y_i = \hat{\theta}_0 + \hat{\theta}_1 x_{i1} + \hat{\theta}_2 x_{i2} + \hat{\theta}_3 x_{i3} i = 1, \dots, n$ and,

PAIRWISE COMPARISON IN REPEATED MEASURES

$f_2 : y_i = \hat{\theta}_0 + \hat{\theta}_1 x_{i1} + \hat{\theta}_2 x_{i2} + \hat{\theta}_3 x_{i3} + \hat{\theta}_4 x_{i4} \ i=1, \dots, n$ offered models, which have been generated from F_{θ_k} . Also we have $g : y_i = 0.3 + 0.5z_{i1} + 3z_{i2} + 1.1z_{i3} + \varepsilon_{i2} \ i=1, \dots, n$

Table 1. Comparison of AIC with A'IC by using from Monte-Carlo simulation for linear regression models f_1, f_2, g_1 , and g_2 .

Size	Model	AIC	A'IC	Δ AIC	Δ A'IC
n=50	f_1	-2990	-2598	-	-
	f_2	-2700	-2210	290	388
	g_1	200	592	3190	2006
	g_2	248	738	3238	1860
n=100	f_1	-3500	-2708	-	-
	f_2	-3200	-2210	300	498
	g_1	430	1222	3930	3930
	g_2	455	1445	3955	4153
n=200	f_1	-5400	-3808	-	-
	f_2	-4360	-2370	1040	1438
	g_1	210	1802	5610	5610
	g_2	240	2230	5640	6038
n=350	f_1	-7230	-4438	-	-
	f_2	-6400	-2910	30	1528
	g_1	325	3117	7555	7555
	g_2	360	3850	7590	8288
n=500	f_1	-9730	-5738	-	-
	f_2	-9300	-4310	430	1428
	g_1	400	4392	10130	10130
	g_2	425	5415	10155	11153

Thus, ε_{i_2} , has been generated as random from distribution $N(0,1)$, and Models $g_1 : y_i = \hat{\beta}_0 + \hat{\beta}_1 z_{i1} + \hat{\beta}_2 z_{i2} + \hat{\beta}_3 z_{i3} \quad i=1, \dots, n$ and $g_2 : y_i = \hat{\beta}_0 + \hat{\beta}_1 z_{i1} + \hat{\beta}_2 z_{i2} + \hat{\beta}_3 z_{i3} + \hat{\beta}_4 z_{i4} \quad i=1, \dots, n$. The models are generated from G_{β_d} . This simulation is achieved by using from software R, and the number of repetitions are 10^3 , and samples $n = 50, 100, 200, 350, 600$, have been considered. The results of simulation are presented in the Table 1.

In the third and fourth columns of Table 1, the value of AIC and A'IC are presented in order to various values of n and for offered models f_1, f_2, g_1 , and g_2 . Therefore the relation between values AIC for offering models is obvious as $AIC(f_1) < AIC(f_2) < AIC(g_1) < AIC(g_2)$.

The family F_{θ_k} is well specified, but the family G_{β_d} is misspecified. Thus, this relation is logical. With attention to the fourth column of Table 1 recent relation also is confirmed for A'IC. In other words $A'IC(f_1) < A'IC(f_2) < A'IC(g_1) < A'IC(g_2)$.

With increasing n , the value of A'IC has been increased for the offered models, but the direction is confirmed unequally. The absolute magnitude difference of the value AIC and A'IC between the model of f_1 and other models is presented in the fifth and sixth columns of table. The absolute magnitude differences have been shown by the symbols of ΔAIC and $\Delta A'IC$. If there are symbols, as

$$\Delta AIC_{|f_1-f_2|} = |AIC(f_1) - AIC(f_2)| \text{ and } \Delta AIC_{|f_1-g_j|} = |AIC(f_1) - AIC(g_j)|, \quad j=1,2$$

$$\Delta A'IC_{|f_1-f_2|} = |A'IC(f_1) - A'IC(f_2)| \text{ and } \Delta A'IC_{|f_1-g_j|} = |A'IC(f_1) - A'IC(g_j)|, \quad j=1,2$$

For $n=50, 100, 150, 200, 350, 500$, and models f_1, f_2, g_1 , and g_2 will be confirmed the relation as

$$\Delta AIC_{|f_1-f_2|} < \Delta AIC_{|f_1-g_1|} < \Delta AIC_{|f_1-g_2|} \text{ and } \Delta A'IC_{|f_1-f_2|} < \Delta A'IC_{|f_1-g_1|} < \Delta A'IC_{|f_1-g_2|}.$$

With attention to these relations the direction of similarity the model selection for information criteria AIC and A'IC for various n have been shown with this the quality that the criterion A'IC is a consistent information criterion.

PAIRWISE COMPARISON IN REPEATED MEASURES

Table 2. Comparison of AIC with A'IC by using Monte-Carlo simulation, for the state that generates model data is Normal standard and offered models are from a Laplace family with different parameters.

Size	Model	AIC	A'IC	Δ AIC	Δ A'IC
n=50	$f_1 = lap(0,1.3)$	-90	106	-	-
	$f_2 = lap(0,1)$	-70	126	20	20
	$f_3 = lap(2,1)$	-56	140	34	34
	$f_4 = lap(-2,0.9)$	-50	146	40	40
n=100	$f_1 = lap(0,1.3)$	-200	196	-	-
	$f_2 = lap(0,1)$	-160	236	40	40
	$f_3 = lap(2,1)$	-143	253	57	57
	$f_4 = lap(-2,0.9)$	-130	266	70	70
n=200	$f_1 = lap(0,1.3)$	-345	451	-	-
	$f_2 = lap(0,1)$	-295	501	50	50
	$f_3 = lap(2,1)$	-255	541	90	90
	$f_4 = lap(-2,0.9)$	-240	556	105	105
n=350	$f_1 = lap(0,1.3)$	-610	786	-	-
	$f_2 = lap(0,1)$	-525	871	85	85
	$f_3 = lap(2,1)$	-487	909	123	123
	$f_4 = lap(-2,0.9)$	-441	955	169	169
n=500	$f_1 = lap(0,1.3)$	-986	1010	-	-
	$f_2 = lap(0,1)$	-865	1131	121	121
	$f_3 = lap(2,1)$	-777	1219	209	209
	$f_4 = lap(-2,0.9)$	-670	11326	316	316

In the third and fourth column Table 2 values of AIC and A'IC for n=50, 100, 200, 350 and 500, have been respectively considered Laplace offered models f_1 , f_2 , f_3 , and f_4 . Therefore the relation between values AIC for offered models of Laplace family is obvious as $A'IC(f_1) < A'IC(f_2) < A'IC(f_3) < A'IC(f_4)$.

With attention to the fourth column in the Table 2, the recent relation is also confirmed for A'IC. In other words, $A'IC(f_1) < A'IC(f_2) < A'IC(f_3) < A'IC(f_4)$. In the fifth and sixth columns the absolute magnitude difference have been presented respectively for the value AIC and A'IC between the model of f_1 and any which from other models to confirm with any n , symbols of ΔAIC and $\Delta A'IC$ has been shown. With attention to these two columns for n 's different have $\Delta AIC = \Delta A'IC$. If we have these symbols as

$$\Delta AIC_{|f_i - f_j|} = |AIC(f_i) - AIC(f_j)| \quad i \neq j \quad \text{and} \quad \Delta A'IC_{|f_i - f_j|} = |A'IC(f_i) - A'IC(f_j)| \quad i \neq j$$

for any $n = 50, 100, 200, 350, 500$, models f_1, f_2, f_3 , and f_4 , confirms the relation as

$$\Delta AIC_{|f_1 - f_2|} < \Delta AIC_{|f_1 - f_3|} < \Delta AIC_{|f_1 - f_4|} \quad \text{and} \quad \Delta A'IC_{|f_1 - f_2|} < \Delta A'IC_{|f_1 - f_3|} < \Delta A'IC_{|f_1 - f_4|}.$$

With attention to these relations, the direction of similarity model selection for information criteria AIC and A'IC for various n has been shown. But the information criterion A'IC is the consistent information criterion.

Conclusion

In this article investigating the inconsistent information criterion AIC, and by eliminating the inconsistency problem, a method for achieving an information criterion has been presented based on Kullback-Leibler risk and the consistent information criterion A'IC has been obtained. Therefore this information criterion is the only consistent information criterion and asymptotically unbiased. It is obtained based on Kullback-Leibler risk. Via simulation for linear regression and classic model, the quality of model selection was shown throughout the two information criterion, AIC and A'IC. According to the consistent information criterion of A'IC, it is possible for further discussion and to refine the other information criteria, which are based on Kullback-Leibler risk (as AICc and KICc) and add the consistency feature to the criteria.

References

- Akaike, H. (1973). Information theory and an extension of maximum likelihood principle. *Second Intention Symposium on Information Theory*, 267-281.
- Bozdogan, H. (2000). Akaike's information criterion and recent developments in information complexity. *Journal of Mathematical Psychology*, 44(1), 62-91.
- Cavanagh, J. E. (1994). A large sample model selection criterion based on kulback's symmetric divergence. *Statistics and Probability Letters*, 44, 333-344.
- Cawley, G., & Talbot, N. (2010). On over-fitting model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research*, 11, 2079-2107.
- Choi, H., & Kiffer, N. M. (2006). Differential geometry and bias correction in nonnested hypothesis testing. Unpublished manuscript, Department of Economics, Cornell University, Ithaca, NY.
- Hall, P. 1987. On Kullbake-Leibler loss and density estimation. *The Annals of Statistics*, 15(4), 1491-1519.
- Hu, B., & Shao, J. (2008). Generalized linear model selection using r. *Journal of Statistical Planning and Information*, 138(12), 3705-3712.
- Konishi, S., & Kitagawa, G. (1996). Generalized information criteria in model selection. *Biometrika*, 83(4), 875-890.
- Kullback, S., & Leibler, R.A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1), 76-86.
- Takeuchi, K. (1976). Distribution of information statistics and criteria for adequacy of models. *Mathematical Sciences*, 153, 12-18.

Optimal Location Design for Prediction of Spatial Correlated Environmental Functional Data

M. Rasekhi

Shahid Chamran University
Ahvaz, Iran

B. Jamshidi

Shahid Chamran University
Ahvaz, Iran

F. Rivaz

Shahid Beheshti University
Tehran, Iran

The optimal choice of sites to make spatial prediction is critical for a better understanding of really spatio-temporal data. It is important to obtain the essential spatio-temporal variability of the process in determining optimal design, because these data tend to exhibit both spatial and temporal variability. Two new methods of prediction for spatially correlated functional data are considered. The first method models spatial dependency by fitting variogram to empirical variogram, similar to ordinary kriging (univariate approach). The second method models spatial dependency by linear model co-regionalization (multivariate approach). The variance of prediction method was chosen as the optimization design criterion. An application to CO concentration forecasting was conducted to examine possible differences between the design and the optimal design without considering temporal structure.

Keywords: Spatio-temporal process, functional data, optimal design, ordinary kriging, total model, GenSA optimization

Introduction

A method for optimum choice of location to obtain better spatial prediction is needed. Ordinary geo-spatial prediction methods deal with scalar value for random variables in each coordinate (Cressie, 1993). Recently, Giraldo et al. (2011a, 2011b) have extended geo-spatial prediction methods for one-dimensional functional data (curves) based on the statistic proposed by Delicado et al. (2010). Delicado et al. constructed a statistic for spatial correlated functional data and proposed a new experimental variogram based on L_2 functional distance to function value spatial data. Giraldo et al. extended the ordinary kriging model. They also

M. Rasekhi and B. Jamshidi are Doctoral Students in the Department of Statistics. Email them at m-rasekhi@phdstu.scu.ac.ir and b-Jamshidi@phdstu.scu.ac.ir, respectively. F. Rivaz is an Assistant Professor in the Department of Statistics. Email her at [F-rivaz@sbu.ac.ir](mailto:rivaz@sbu.ac.ir).

used the total linear functional model for creating a new prediction method. In the next section, these models are introduced briefly.

Many researchers have investigated the problem of spatial sampling design e.g. Zhu & Stein (1999), Wikle & Royle (1999), Diggle & Lophaven (2006), Fuentes et al. (2007), etc. Here, variance of spatial prediction in extended ordinary kriging and Total model spatial prediction of function value data are chosen as criteria for obtained optimization. This is done by application of Xiang et al.'s (2012) optimization procedure, “Generalized Simulated Annealing for Efficient Global Optimization” or “GenSA,” described in the section titled ‘Optimization Procedure.’ Following that, CO concentration data in Tehran weather pollution stations is introduced as spatial correlated functional data in seventeen stations. The proposed approach is illustrated in the section titled ‘Application,’ and for the possible differences between this design and the optimal design without considering temporal structure, the air monitoring network is redesigned based on the average data over time.

Prediction Procedures

Consider a functional spatial process $X = \{X_s(t) : s \in D \subseteq R^d, t \in [a, b]\}$, where functional variable X_s belongs to the separable Hilbert space H of square integrable functions defined on T for any $s \in D$. We assume second-order stationarity and isotropicity for each $t \in T$ in random process. Let s_1, \dots, s_n be the sites in D that we observe a realization X_{s_1}, \dots, X_{s_n} of the functional random process X_s .

Ordinary kriging for function value spatial data

Ordinary kriging to function value spatial data is extended by Giraldo e.al (2011a) as following model

$$\hat{X}_{s_0}(t) = \sum_{i=1}^n \lambda_i X_{s_i}(t); \quad \lambda_1, \dots, \lambda_n \in R, t \in T \quad (1)$$

where $\hat{X}_{s_0}(t)$ is predicted function in location s_0 . Modeling based on this prediction method needs some assumptions such as functional version of intrinsic stationarity and isotropicity (see Delicado et al., 2010). Empirical variogram for modeling spatial dependency is obtained founded on the following minimization variance considering unbiasedness condition

$$\min_{\lambda_1, \dots, \lambda_n} \int_T V(\hat{X}_{s_0}(t) - X_{s_0}(t)) dt \text{ s.t. } \sum_{i=1}^n \lambda_i = 1 \quad (2)$$

where moment estimation method leads to

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{i,j \in N(h)} \int_T (\chi_{s_i}(t) - \chi_{s_j}(t))^2 dt \quad (3)$$

Therefore, ordinary variogram, including exponential, spherical and so on, is fit to (3) by least square method and scale value coefficients are obtained.

Total model prediction of function value spatial data

Giraldo et al. (2011b) create new predictor for function value spatial data based on functional total model (Ramsay & Silverman, 2005)

$$Y_i(t) = \beta_0(t) + \int_T X_i(v) \beta(v, t) dv + \varepsilon_i(t) \quad (4)$$

and multivariate spatial predictor (Ver Hoef & Cressi, 1993; Ver Hoef and Bari, 1998)

$$\begin{pmatrix} \hat{X}_1(s_0) \\ \vdots \\ \hat{X}_p(s_0) \end{pmatrix} = \begin{pmatrix} \lambda_{11}^1 & \dots & \lambda_{1p}^1 & \dots & \lambda_{n1}^1 & \dots & \lambda_{np}^1 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \lambda_{11}^p & \dots & \lambda_{1p}^p & \dots & \lambda_{n1}^p & \dots & \lambda_{np}^p \end{pmatrix} \begin{pmatrix} X_1(s_1) \\ \vdots \\ X_p(s_1) \\ \vdots \\ X_1(s_n) \\ \vdots \\ X_p(s_n) \end{pmatrix} \quad (5)$$

as follows

$$\hat{X}_{s_0}(v) = \sum_{i=1}^n \int_T \lambda_i(t, v) X_{s_i}(t) dt, \quad v \in T \quad (6)$$

where $\lambda_n(t, v)$, ..., $\lambda_1(t, v)$ are functions $T \times T \rightarrow R$. Modeling based on this prediction need some assumption such as functional version of second-order stationarity and isotropicity that is proposed by Delicado et al. (2010). Coefficients of this model are found through following minimization variance with unbiasedness condition

$$\min_{\lambda_1(\cdot, \cdot), \dots, \lambda_n(\cdot, \cdot)} \int_T V(\hat{\chi}_{s_0}(v) - \chi_{s_0}(v)) dv \text{ s.t. } E(\hat{\chi}_{s_0}(v)) = E(\chi_{s_0}(v)) \quad (7)$$

To solve equation (7), spatial dependency must be modeled with linear model co-regionalization (Wackernagel, 2003).

Optimization Procedure

Optimization is the process which one finds that value of a vector x , say, that maximizes or minimizes a given function f . The idea of optimization goes to the heart of statistical methodology, as it is involved in solving statistical problems based on least squares, maximum likelihood, posterior mode, and so on. Xiang et al. (2012) created global optimization procedure “GenSA” which is applicable for geo-statistical process. GenSA gives the lower and upper bound of geographical coordinates and finds optimum coordinates based on specific criteria. GenSA uses a distorted Cauchy-Lorentz visiting distribution, with its shape controlled by the parameter q_v

$$g_{q_v}(\Delta x(t)) \propto \frac{(T_{q_v}(t))^{\frac{D}{3-q_v}}}{\left(1 + (q_v - 1) \frac{(\Delta x(t))^2}{(T_{q_v}(t))^{\frac{2}{3-q_v}}}\right)^{\frac{1}{q_v-1} + \frac{D-1}{2}}} \quad (8)$$

Here t is the artificial time. This visiting distribution is used to generated a trial jump distance $\Delta x(t)$ of variable $x(t)$ under artificial temperature $T_{q_v}(t)$. The trial jump is accepted if it is downhill (in terms of the objective function). If the jump is uphill it might be accepted according to an acceptance probability. A generalized Metropolis algorithm is used for the acceptance probability

$$p_{q_v} = \min \left\{ 1, (1 - q_a) \beta \Delta E \right\}^{\frac{1}{1-q_a}} \quad (9)$$

where q_a is a parameter.

The minimax of prediction model of variance was used as optimization criterion. Variance of mentioned predictions is calculated as follows

$$\sigma_{s_0}^2 = \int_T V(\hat{\chi}_{s_0}(t) - \chi_{s_0}(t)) dt \quad (10)$$

Giraldo et al. (2011a) calculate variance of ordinary kriging, resulting in

$$\sigma_{s_0}^2 = \sum_{i=1}^n \lambda_i \int_T \Gamma_{s_i, s_0}(t) dt - \mu = \sum_{i=1}^n \lambda_i \gamma(h_i) - \mu \quad (11)$$

where functional variogram $\Gamma(t)$ is defined by Delicado et al. (2010) and $\gamma(\cdot)$ is the classical variogram fitted to empirical variogram. Giraldo et al. (2011b) calculate variance of Total model, resulting in

$$\sigma_{s_0}^2 = \int_T V(\hat{\chi}_{s_0}(v) - \chi_{s_0}(v)) dv = \sum_{i=1}^n Tr(\hat{C}_i^T Q_{ii} \hat{C}_i W) + 2 \sum_{i < j} Tr(\hat{C}_i^T Q_{ij} \hat{C}_j W) \quad (12)$$

More detail of this variance is provided in Giraldo et al. (2011a, 2011b). To calculate the predictions variance in any location s_0 , a smoothing process is applied to the curves, which expands the curves and the functional parameters in terms of a set of Fourier basis functions. The number of Fourier basis is found by a Functional Cross-Validation procedure similar to the leave-one-out procedure (Cressie, 1993) introduced by Giraldo et al. (2011a).

Data

Transportation-related air pollution is one of Tehran's most important problems. One of the most hazardous air polluting agents is carbon monoxide (CO), often exceeding two or three times the average levels recommended by the World Health Organization (WHO). This gas is colorless, odorless, and tasteless, and its predilection to bind hemoglobin is 200-220 times more than that of oxygen. Thus, it can prevent oxygen transfer to tissues and cause tissue hypoxia. For these reasons, the demand for reliable data to assess progress in air quality has grown rapidly over the past decade. In fact, motivated by increasing air monitoring stations in Tehran, three newly designed sites are proposed for monitoring CO. The data set used here describes daily averages of carbon monoxide (in ppm) at 17 monitoring sites,

OPTIMAL LOCATION DESIGN FOR PREDICTION OF DATA

geographically distributed across Tehran. Following the air quality standards (NAAQS, <http://www.epa.gov/air/criteria.html>), daily CO concentration is measured as the daily 1-hour average concentration. The current analysis considers data collected in 2011.

Application

Positioning of Tehran air quality stations and curves of seventeen stations created by smoothing procedure (Ramsay & Silverman, 2005) with 31 Fourier basis is illustrated in Figure 1.

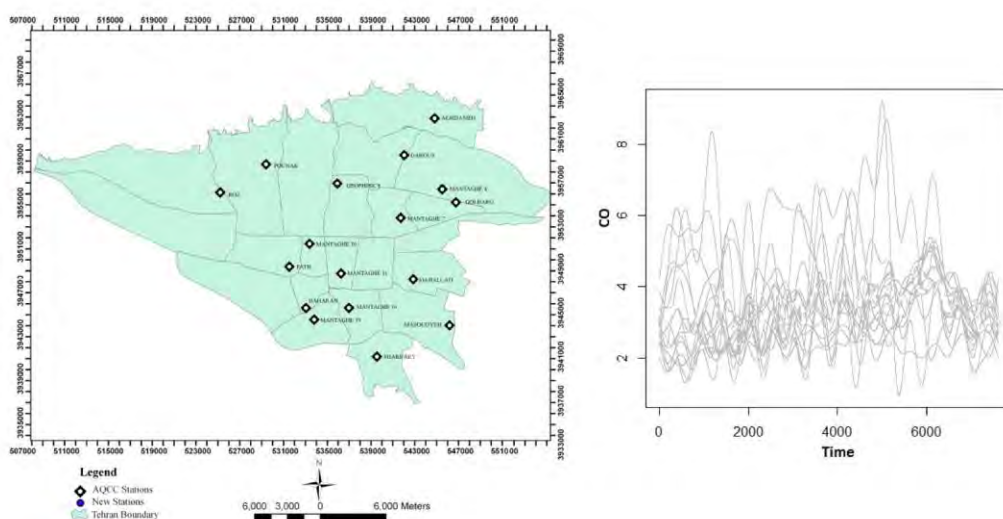


Figure 1. Position and CO concentration curves of seventeen air quality monitoring stations in Tehran

Considering variance of the mentioned prediction method in GenSA algorithm, three locations in Tehran map can be identified that minimize maximum predictions variance simultaneously. The possible differences between this design and the optimal design, without considering temporal structure, are surveyed, and the air monitoring network is redesigned based on averaged data over time. Figure 2 shows the optimal location based on Ordinary method for functional data vs. ordinary kriging on averaged time data (spatial data). It is worth mentioning that spherical

variogram (13) is chosen for modeling dependency structure, then parameters are estimated based on empirical variogram by applying least square method.

$$\gamma(\|h\|) = \begin{cases} \tau^2 + \sigma^2 & \|h\| > 1/\varphi \\ \tau^2 + \sigma^2 \left(\frac{3\varphi\|h\|}{2} - \frac{1}{2}(\varphi\|h\|)^3 \right) & \|h\| < 1/\varphi \\ 0 & o.w \end{cases} \quad (13)$$

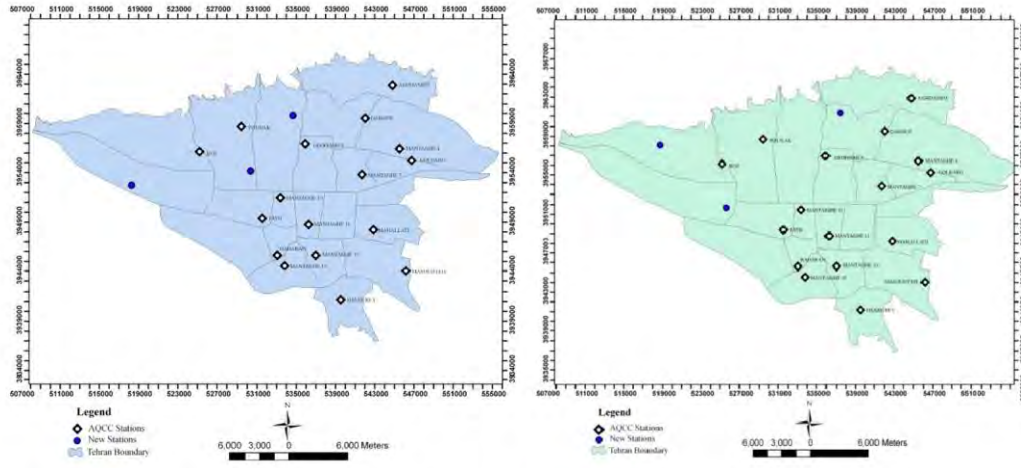


Figure 2. Optimal location for three monitoring sites based on kriging model for spatial data (left) and ordinary model for spatial functional data (right)

Figure 3 shows optimal location founded on Total prediction model for functional data vs. ordinary kriging on averaged time data (spatial data). Dependency structure is modeled with a linear model co-regionalization with exponential variogram (14)

$$\gamma(\|h\|) = \begin{cases} \tau^2 + \sigma^2 (1 - \exp(-\varphi\|h\|)) & \|h\| > 0 \\ 0 & o.w \end{cases} \quad (14)$$

OPTIMAL LOCATION DESIGN FOR PREDICTION OF DATA

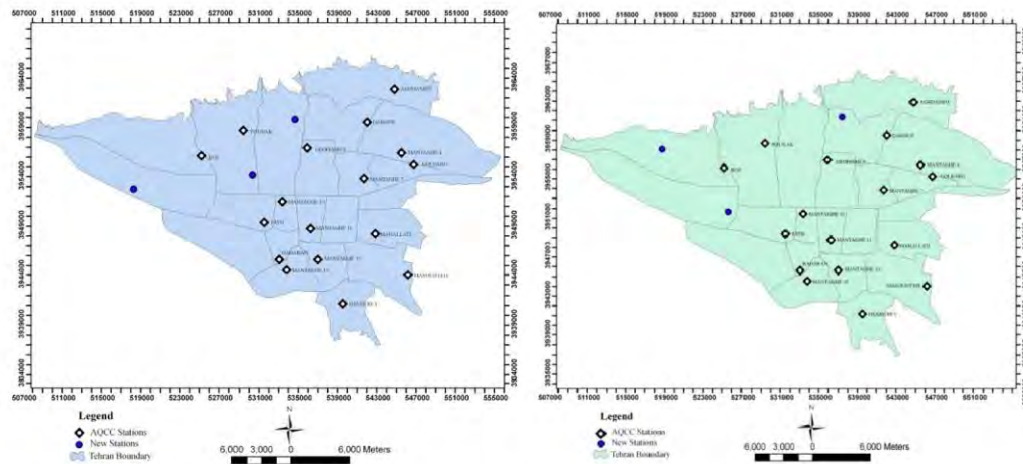


Figure 3. Optimal location for three monitoring sites based on kriging model for spatial data (left) and Total model for spatial functional data (right)

Results

Figures 2 and 3 show the optimal design based on ordinary model and Total model for the spatially correlated functional data.

Conclusion

Although the modeling of spatial dependency based on the two proposed functional models is different, both tend to locate new monitoring stations nearer from existing stations than in the non-functional version of ordinary kriging. Thus, considering time on spatial data affects location sampling. In other words, maximum variance of functional predictions of the three locations is global minimized closer to other stations, but optimal design based on averaged data over time (spatial data) tends to fill the space (Figures 2 and 3).

References

- Cressie, N. (1993). *Statistics for spatial data*. New York: John Wiley and Sons.
- Delicado, P., Giraldo, R., Comas, C., & Mateu, J. (2010). Statistics for spatial functional data: Some recent contributions. *Environmetrics*, 21, 224-239.

- Diggle, P., & Lophaven, S. (2006). Bayesian geostatistical design. *Scandinavian Journal of Statistics*, 33, 53-64.
- Fuentes, M., Chaudhuri, A., & Holland, D. M. (2007). Bayesian entropy for spatial sampling design of environmental data. *Environmental and Ecological Statistics*, 14, 323-340.
- Giraldo, R., Delicado, P., & Mateu, J. (2011a). Ordinary kriging for function-value spatial data. *Journal of Environmental and Ecological Statistics*, 18, 411-426.
- Giraldo, R., Delicado, P., & Mateu, J. (2011b). Geostatistics with infinite dimensional data: A generalization of cokriging and multivariable spatial prediction. *Matemática: ICM-ESPOL*, 9, 16-21.
- Ramsay, J., & Silverman, B. (2005). *Functional data analysis* (2nd ed.). Berlin: Springer.
- Ver Hoef, J., & Barry, R. (1998). Constructing and fitting models for cokriging and multivariable spatial prediction. *Journal of Statistical Planning and Inference*, 69, 275-294.
- Ver Hoef, J., & Cressie, N. (1993). Multivariable spatial prediction. *Mathematical Geology*, 25(2), 219-240.
- Wackernagel, H. (2003). *Multivariate Geostatistics: An Introduction With Applications* (3rd ed.). Berlin: Springer.
- Wikle, C. K., & Royle, J. A. (1999). Space-Time Dynamic Design of Environmental Monitoring Networks. *Journal of Agricultural, Biological, and Environmental Statistics*, 4, 489-507.
- Xiang, Y., Gubian, S., Suomela, B., & Hoeng, J. (2013). Generalized Simulated Annealing for Global Optimization: The GenSA Package. *The R Journal*, 5(1), 13-29. Retrieved at <http://journal.r-project.org/archive/2013-1/xiang-gubian-suomela-et-al.pdf>
- Zhu, Z., & Stein, M. L. (1999). Spatial sampling design for parameter estimation of the covariance function. *Journal of Statistical Planning and Inference*, 134, 583-603.

Missing Data and the Statistical Modeling of Adolescent Pregnancy

Dudley L. Poston
Texas A&M University
College Station, TX

Eugenia Conde
Rutgers University
New Brunswick, NJ

Missing data is a pervasive problem in social science research. Many techniques have been developed to handle the problem. Different ways of handling missing data were shown to lead to different results in statistical models. A demonstration was given based on statistical modeling of the likelihood of a woman reporting having had an adolescent pregnancy by handling missing data with several different approaches. Results indicate that many of the independent variables in the model vary in whether they are, or are not, statistically significant in predicting the log odds of a woman having a teen pregnancy, and in the ranking of the magnitude of their relative effects on the outcome.

Keywords: Missing data, listwise deletion, mean substitution, multiple imputation, proxy variables, adolescent pregnancy, race/ethnicity, logistic regression, logit coefficients, semi-standardized logit coefficients, demography

Introduction

Missing data is a pervasive problem in social science research. “Sooner or later, usually sooner, anyone who does statistical analysis runs into problems with missing data” (Allison, 2001: 1). Many techniques have been developed to handle missing data; often, the results of a statistical model will differ depending on the technique used.

Missing Data Mechanisms

According to Rubin (1976; 1987), there are three missing data mechanisms; the data are either “missing completely at random” (MCAR), “missing at random” (MAR) or “missing not at random” (MNAR). Missing data are said to be missing completely at random (MCAR) when the probability of the missing data for a

Dr. Poston is Professor of Sociology. Email him at d-poston@tamu.edu. Eugenia Conde is a Postdoctoral Fellow.

variable does not depend on the variable itself or on any of the other independent variables in the model. MCAR refers to the “condition in which missing responses to a particular variable are independent of the values of any other variable in the explanatory model and of the true value of the variable in question” Treiman (2009, p. 182). If all the missing data are MCAR, this is usually not a serious problem because the remaining data are considered to be a subsample of the original sample.

Missing data are considered to be missing at random (MAR) if the probability of the missing data does not depend on the values of variables with the missing data, after controlling for other variables in the model. That is, MAR refers to “the condition in which missingness is independent of the true value of the variable in question but not of at least some of the other variables in the explanatory model” (Treiman, 2009, p. 182).

Missing data are considered to be missing not at random (MNAR) when the MAR assumption is violated. The data are MNAR if the probability that the values were missing depends on the variable itself.

Methods for Handling Missing Data

There are many methods for handling missing data. We discuss several of the more popular approaches and then use each separately in an analysis of adolescent pregnancy.

1. Listwise Deletion The method that is the default method in most statistical packages is listwise deletion, also known as case deletion. It drops the missing values from the data set, and the analysis is then conducted using the reduced sample. If the data are MCAR, the resulting smaller sample is considered to be an unbiased subsample of the original dataset (Allison, 2001), and the use of listwise deletion should result in models with unbiased estimates. However, the standard errors will be slightly larger because the sample size is now, obviously, smaller. Statistical power will be reduced and the probability of finding significant results decreased; thus the listwise deletion method is often viewed as conservative provided that the MCAR assumption has been met (Acock, 2005). But if the missing data are MAR and listwise deletion is used, then the estimates will likely be biased (Allison, 2001).

2. Mean Substitution Mean substitution is a very simple approach. The missing values for a variable are replaced with the mean value for that variable. Mean substitution is especially problematic when the percentage of missing values

is large because this greatly reduces the variance and hence underestimates the correlation between the variable with missing values and any of the other variables in the model (Acock, 2005; Allison, 2001). Mean substitution “is possibly the worst missing data handling method available” Enders (2010, p. 43).

3. Mean Substitution for Subgroups A modification of mean substitution assigns the mean values for subgroups of the analysis. For example, a researcher might handle missing data on a variable such as income for the males and females in the sample by assigning to the males the average value of income for males, and to the females the average value of income for females. Although this modification reduces the variance, it is considered to be only slightly better than substituting with the overall mean (Acock, 2005).

4. Proxy Method When confronted with an excessive amount of missing data on an independent variable, some have used the proxy method as a solution. That is, they have substituted for the variable with the missing data another variable with little or no missing data that is related substantively and statistically to the variable with the missing data. For example, to address the situation of an excessive amount of missing data on a variable such as income, one could use educational attainment as a proxy for income.

5. Dropping the Variable(s) with Missing Data This approach simply drops from the analysis the variable (or variables) with excessive amounts of missing. It should be avoided without question because of the obvious problem of model misspecification.

The above are five of the “traditional” methods used for handling missing data. With the exception of listwise deletion when the data are MCAR, all five are problematic. For one thing, they will often produce biased estimates and inefficient standard errors. And when listwise deletion is used with MAR data, the estimates will be biased and the standard errors inefficient.

(Other traditional methods not used in this paper include dummy variable adjustment and hot and cold deck imputation. Dummy variable adjustment uses all the cases and adjusts for those that have missing values by adding a dummy variable scored 1 if the value for the variable is missing, and 0 if not missing. Hot deck imputation also uses all the cases but replaces the missing values with random values found in the observed data. Cold deck imputation is similar but replaces the missing values with those from another data set. These methods may seem to be appealing because they use all the cases, but they have been shown to produce

biased estimates irrespective of whether or not the data are MCAR, MAR or MNAR (Acock, 2005; Allison, 2001).

6-8. Multiple Imputation (MI) - three versions The most popular of the non-traditional methods is multiple imputation (MI), a method first introduced by Rubin in 1987. There are several variations of MI.

It has been argued that MI is the preferred method for handle missing data because “when used correctly, it produces estimates that are consistent, asymptotically efficient and asymptotically normal when the data are MAR” (e.g., Allison, 2001, p. 27). MI has become the gold-standard approach for dealing with missing data (Treiman, 2009, p. 186-186).

Multiple imputation is not concerned with recovering the missing data like the traditional methods mentioned above. Instead, it is concerned with estimating the population variances so as to produce generalizable estimates (Acock, 2005; Allison, 2001; Enders, 2010; Rubin, 1987). Unique about this method is that it does not treat the data as if “they were real” (Allison, 2001). Instead MI estimates the values by taking into account the uncertainty of the missing values. MI recognizes that even if the missing values are imputed, there is still uncertainty in those values, so it adjusts the variances to take this into account.

MI has three steps: imputation, analysis, and the combination of datasets. The imputation stage creates several data sets; the analysis stage runs the desired analysis in each data set; and the combination stage combines the results from the imputations using rules developed by its creator, Donald Rubin.

In the imputation stage, auxiliary variables may or may not be used to impute the missing values. Auxiliary variables are used that are statistically related to the variables with missing values, so to enhance the effectiveness of the imputation stage. The auxiliary variables are not used as independent variables in the regression equation per se, but are used to provide more information about the variances of the independent variables with the missing data. A preferred MI equation is usually one that uses auxiliary variables (Allison, 2001; Treiman, 2009).

The two main MI iterative methods for handling missing data are the fully conditional specification (FCS) method, and the Markov chain Monte Carlo (MCMC) method. The fully conditional specification (FCS) method is sometimes known as imputation by chain equation (ICE); it imputes continuous and categorical variables without assuming a multivariate normal distribution. Simulation studies have shown that it works reasonably well, and the results are comparable to the MCMC method (Lee & Carlin, 2010).

The Markov chain Monte Carlo (MCMC) method is an iterative procedure that assumes a multivariate normal distribution of all the variables in the model. It works best when imputing continuous variables (Schafer, 1997), but it can also be used to impute categorical variables (Allison, 2001; Lee and Carlin, 2010).

Following the above discussion, we will use three MI methods in our analysis of adolescent pregnancy, as follows: 6. MI using the fully conditional specification (FCS) method; 7. MI using the Markov chain Monte Carlo (MCMC) method with auxiliary variables; and 8. MI using the Markov chain Monte Carlo (MCMC) method but only imputing education and income.

Thus, eight models of adolescent pregnancy will be estimated, with missing data handled differently in each of the eight models.

Data and Method

Data were taken from the National Longitudinal Study of Adolescent Health (Add Health) (Harris, 2008), a nationally representative stratified sample of adolescents in the 7th through the 12th grades who were followed across four waves between 1994 and 2008. The sample was collected from 80 high schools and 52 middle schools and junior high schools across the United States. The first wave of data was collected in 1994-1995, the second in 1996, the third in 2001-2002, and the fourth in 2007-2008. Data on the parents of the school children were collected in the first wave. We use data from wave I and wave III for the female students and their parents.

Logistic regression is used to estimate the log odds of females who had a pregnancy when they were between the ages of 15-19. Seven theoretically relevant independent variables were selected, as follows: (1) a dummy variable from wave 1 regarding whether or not the adolescent ever made a pledge to remain a virgin until marriage, scored 1 if yes and 0 if no; (2) the adolescent's race/ethnicity measured with a series of dummy variables (African American, non-Hispanic white, Mexican-origin, other Latina; other race; and non-Hispanic white, which was used as the reference); (3) the adolescent's religion measured with six dummy variables (no religion, Protestant, Evangelical Protestant, Black Protestant, other religion, and Catholic; the Catholic dummy was used as the reference group); (4) household income as reported by the parent in wave 1 (measured in thousands) with \$100,000 as the ceiling; (5) parental education as reported by the parent in wave 1 and measured as number of years of school completed; (6) the importance of religion to the adolescent ("How important is religion to you?"), ranging from a value of 1 if the woman reported no religious affiliation or responded "not important at all" to

a value of 4 if she reported “very important”; and (7) the adolescent’s perceived likelihood to attend college, with 1 as the lowest category and 5 as the highest. All these independent variables have been previously shown to be influential in models predicting whether or not a woman had a teen pregnancy (see, e.g., [Bean and Swicegood 1985](#); [Klepinger et al., 1995](#); [Rosenbaum, 2006](#)).

Results

Table 1. Descriptive Data: 6,719 Females, The National Longitudinal Study of Adolescent Health, Waves 1 and 3

Variable	Cases	Percent missing	Mean	SD
Dependent Variable				
Teen pregnancy	6,710	0.24	0.18	0.38
Seven Independent Variables				
1. Virginity pledge	6,644	1.22	0.15	0.36
2. Race / Ethnicity	6,719	0.10		
White	3,568		0.67	0.47
African American	1,510		0.17	0.37
Mexican	539		0.06	0.24
Other Latina	538		0.05	0.23
Other	564		0.05	0.21
3. Religion	6,620	1.60		
Catholic	1,757		0.24	0.43
None	744		0.12	0.32
Protestant	1,447		0.22	0.42
Evangelical	1,056		0.20	0.40
Black Protestant	884		0.11	0.31
Other	682		0.11	0.31
Jewish	50		0.01	0.09
4. Household Income (in thousands)	4,983	26.00	42.70	27.00
5. Parental Education (in years)	5,708	15.14	13.27	2.45
6. Religious importance	6,717	0.13	3.12	0.93
7. Likelihood of college	6,681	0.67	4.25	1.13

Table 1 presents descriptive data on the dependent variable and the independent variables for the 6,719 females of age 20 years or higher in our sample. We show in the first data column the number of women for whom we have data for each variable. The maximum number of cases is 6,719. In column 2 we show the percentage of the cases with data missing for each variable. Of the nine variables

MISSING DATA AND THE STATISTICAL MODELING OF ADOLESCENT

we use in the logit regression equations (the dependent variable and eight independent variables), only three have missing data percentages of more than one percent: household income, 26.0 percent; parental education, 15.1 percent; and religion 1.6 percent. With more than one quarter of the cases having missing data on income, this means we would lose at least this percentage of respondents from the analysis were we to rely on listwise deletion as the method for handling missing data.

In the third data column of [Table 1](#), note that 18 percent of the women in the sample reported having had a teen pregnancy, 15 percent reported having made a pledge while a teenager to remain a virgin until marriage. Almost 67 percent of the respondents were white, and their mean household income was over \$42.7 thousand. Religion was fairly to very important for most of the respondents, and most of them believed it is very likely that they will attend college.

These data were analyzed using the eight different approaches discussed above for handling missing data:

1. Listwise deletion
2. Overall mean substitution
3. Mean substitution where the mean values were substituted on the basis of the race and ethnic groups of the women
4. The proxy method where mother's education was used as a proxy for income
5. Dropping the variables with excessive amounts of missing data; parental education and household income, the two variables with the most missing data, were excluded from the equation
6. Multiple imputation in which we imputed all the variables with missing data using the fully conditional specification iterative method
7. Multiple imputation using the Markov chain Monte Carlo iterative method with four auxiliary variables (via four auxiliary variables: Two questions were asked of the parents, namely, "How important is religion to you?" and "Do you have enough money to pay your bills." And two questions were asked of the students, namely, "Since school started this year, how often do

you have trouble getting along with your teachers?” and “How much do you want to go to college?” All four auxiliary questions were answered on a 1-4 or a 1-5 point scale from low to high)

8. Multiple imputation using the Markov chain Monte Carlo iterative method to impute only the two variables with the most missing data, namely household income and parent education. In each of these three MI applications, a total of 100 imputations were undertaken. The 16 cases (only 0.2 percent of all the respondents) that were missing in the teen pregnancy dependent variable were imputed in the imputation stage, but they were dropped from the analysis (von Hippel, 2007).

Because the Add Health Survey is based on multistage probability sampling, one cannot make inferences with these data to the larger population of U.S. women from which the sample was drawn without first taking into account the sampling design. Thus, the “svy” suite of statistical sample adjustment methods available in the Stata 12 statistical package (StataCorp, 2011) was used to introduce survey adjustment estimators.

The results from eight logistic regressions modelling the log likelihood of a woman becoming pregnant while a teenager are compiled in Table 2. Each regression equation handles missing data in a different way, as discussed earlier. The preferred method for handling missing data is multiple imputation using auxiliary variables, shown as model 7 (M7) in the table.

The values in the first line for each variable in Table 2 are the logistic regression coefficients predicting the log odds of a woman having an adolescent pregnancy; if the coefficient is statistically significant, it is asterisked (see legend at the bottom of the table). Immediately below the logit coefficient is its semi-standardized coefficient; this is the logit coefficient that has been standardized in terms of the variance of the independent variable, that is, the logit coefficient has been multiplied by its standard deviation (Long & Freese, 2006, p. 96-98). Alongside each of the semi-standardized coefficients that is statistically significant, in parentheses, is shown the ranking in that equation of its relative effect on the outcome of teen pregnancy.

MISSING DATA AND THE STATISTICAL MODELING OF ADOLESCENT

Table 2. Eight Logistic Regression Models of Teen Pregnancy According to the Method Used to Handle Missing Data: Females Surveyed in The National Longitudinal Study of Adolescent Health, Waves 1 and 3

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
1. Virg-Pledge	.455** -.164(4)	-.323* -.117(8)	-.322* -.117(7)	-.420** .152(5)	-.307* -.112(8)	-.331* -.119(5)	-.328* -.118(5)	-.327* -.118(7)
2. Race/ethnicity								
White	Ref	ref	ref	ref	ref	ref	ref	Ref
African American	.351† .125(6)	.369* .137(6)	.340* .126(6)	.507*** .184(3)	.485*** .180(3)	.232 .082	.249 .088	.343* .122(6)
Mexican-origin	.602* .135(5)	.535* .127(7)	.493† .116(8)	.591* .137(7)	.691** .163(5)	.394 .088	.401 .090	.482† .108(8)
Other Latina	.245 .056	.325† .073(9)	.296 .066	.360* .083(9)	.462** .104(6)	.247 .057	.253 .058	.295 .068
Other	-.035 -.007	-.145 -.031	-.141 -.030	.072 .015	-.081 -.017	-.174 -.035	-.170 -.034	-.157 -.031
3. Religion								
Catholic	Ref	ref	ref	ref	ref	ref	ref	Ref
None	.148 .048	.038 .012	.038 .012	.176 .057	.089 .029	-.008 -.003	.035 .011	.026 .008
Protestant	.254 .108	.183 .076	.183 .076	.228 .095	.175 .073	.189 .080	.191 .081	.185 .079
Evangelical	.306 .121	.365* .145(4)	.368* .146(4)	.449** .178(4)	.459** .182(4)	.344* .136(4)	.343* .135(4)	.351* .138(5)
Black Protestant	.757*** .220(2)	.726*** .224(1)	.722*** .223(1)	.763*** .230(1)	.766*** .236(2)	.748*** .217(2)	.711*** .206(3)	.699*** .203(2)
Other	.162 .050	.148 .046	.149 .046	.262 .081	.190 .059	.133 .042	.133 .042	.143 .045
Jewish	-.258 -.023	-.831 -.079	-.828 -.078	-.771 -.076	-1.021 -.097	-.761 -.069	-.745 -.068	-.757 -.069
4. Hh Income	-.010*** -.281(1)	-.009*** -.217(2)	-.009*** -.220(2)			-.010*** -.259(1)	-.009*** -.253(1)	-.009*** -.254(1)
5. Par-Educ	-.016 -.039	-.023 -.052	-.021 -.048	-.057** -.139(6)		-.024 -.059	-.023 -.056	-.021 -.050
6. Relig-imp	-.127† -.113(7)	-.158* -.143(5)	-.158* -.144(5)	-.119† -.106(8)	-.157* -.142(7)	-.130* -.116(6)	-.115† -.102(6)	-.162* -.144(4)
7. College Lik	-.172*** -.188(3)	-.190*** -.211(3)	-.190*** -.211(3)	-.200*** -.218(2)	-.229*** -.254(1)	-.191*** -.208(3)	-.191*** -.209(2)	-.175*** -.191(3)
Intercept	-.129	.142	.129	.021	-.449	.119	.044	.075
F	6.82	8.06	7.84	8.01	8.55	8.59	8.06	9.27
N	4,822	6,530	6,530	5,557	6,530	6,710	6,710	6,530

†p<0.05 (one tail); *p<0.05 (two tail); **p<0.01 (two tail); ***p<.001 (two tail)

Table 2 (contd.)**Model 1:** Listwise deletion**Model 2:** Full Mean substitution**Model 3:** Mean substitution by race and ethnicity**Model 4:** Education as a proxy for income**Model 5:** Income and education variables dropped**Model 6:** Multiple imputation using the fully conditional specification method**Model 7:** Multiple imputation using Markov chain Monte Carlo method with auxiliary variables**Model 8:** Multiple imputation using Markov chain Monte Carlo method (imputed only education and income)

The regression results in Table 2 indicate that for some independent variables, whether they are or are not statistically significant does not depend at all on which missing data method is used. The virginity pledge variable is statistically significant in predicting the likelihood of a woman having an adolescent pregnancy in all eight equations, as are the Black Protestant variable, the household income variable, the importance of religion variable, and the likelihood to attend college variable. Five variables are not statistically significant in any of the eight equations, namely, Other race/ethnicity, No religion, Protestant religion, Other religion, and Jewish religion.

However, the statistical significance of all the other variables depends on which missing data method is used in the equation. In the preferred equation, Model 7 (see above), being an African American has no significant effect on the likelihood of having an adolescent pregnancy; but it does have an effect on adolescent pregnancy in six of the other equations. The same pattern holds for the Mexican origin variable and for the Other Latina variable.

A woman being an Evangelical does not have a statistically significant effect on the likelihood of her having an adolescent pregnancy if listwise deletion (M1) is used as the method for handling missing data. But being an Evangelical does have a significant effect on the outcome in all seven of the other equations. Similarly parental education has a significant effect on the outcome in the equation where it is used as a proxy for income (M4), but it does not have a significant effect in any of the other equations.

Clearly, for many of the variables, the method used to handle missing data has an important influence on whether or not the independent variables have significant effects in models of adolescent pregnancy. The statistical significance of most of the race/ethnicity variables (African American, Mexican-origin, Other Latina) depends on the method used for handling missing data; if certain methods are used, e.g., mean imputation, these variables are significant in predicting the outcome; if other methods are used, e.g., two of the three multiple imputation

methods, including the preferred method (M7), these variables are not significant. A similar statement may be made regarding one of the religion variables (Evangelical) and the parental education variable.

Another way to evaluate the logit regression results in Table 2 is via the rankings of the statistically significant semi-standardized coefficients. As noted above, these are the logit coefficients that have been standardized in terms of the variances of their independent variables, that is, the logit coefficients are multiplied by their standard deviations (Long & Freese, 2006, p. 96-98). Although there is a problem in the interpretation of the meaning of a semi-standardized coefficient when the independent variable is a dummy variable (there are many dummy variables in the equations, Long, 1997; Poston, 2002, p. 342), their values nonetheless indicate the relative effects of each of the independent variables on the log odds of the woman having a teen pregnancy. In the second row for each variable in each of the eight columns of Table 2 we show the rankings of the magnitude of the semi-standardized coefficient in predicting the outcome. In four of the equations, household income is ranked first, that is, in four equations it has the greatest relative effect on the outcome of adolescent pregnancy; but in two of the equations, those using mean substitution (M2 and M3), it has the second greatest relative effect.

The degree the virginity pledge is influential in predicting the outcome varies according to the method used to handle missing data. If listwise deletion (M1) is used, this variable has the 4th most influential effect, but if mean substitution (M2) is used it has the 8th most influential effect on the outcome. The importance of the effect on the outcome of a woman being an African American varies from the 3rd most important effect in two of the equations (M4 and M5) to the 6th most important effect in four of the equations (M1, M2, M3 and M8). The relative effect on the outcome of the importance of religion variable varies from the 4th most important effect in one equation (M8) to the 8th most important effect in another equation (M4). Clearly the importance of the relative effects of the independent variables on the likelihood of a woman having an adolescent pregnancy vary considerably depending on how missing data are handled in the regression equation.

Discussion

The results show that the levels of significance of the effects, the size of the effects, and their relative importance vary considerably depending on the method used to handle the missing data. Understanding differences between minority group members and whites, and the differential influences of minority membership on an outcome such as adolescent pregnancy is a very important sociological question

with substantial political and social implications. But the issue of how a researcher chooses to handle the missing data can have an impact on how this social issue is understood. If a researcher used listwise deletion or mean substitution to handle the problem of missing data in equations modelling whether or not a woman had an adolescent pregnancy, the conclusion would be after controlling for all the other variables in the model, Mexican origin women and African American women were more likely than White women to have had an adolescent pregnancy. But if multiple imputation with auxiliary variables as the method to handle the missing data, the results would indicate no statistically significant difference between Mexican origin women and African American women compared to White women with regard to the odds of having had a teen pregnancy. In other words, listwise deletion, the default method in most statistical packages, and multiple imputation with auxiliary variables, the so-called “gold standard,” gave the opposite results regarding the odds of a minority woman as compared with a White woman having an adolescent pregnancy.

After controlling for other relevant variables, are minority women more likely than white women to have had an adolescent pregnancy? If listwise deletion or mean substitution was used to handle missing data, the answer is yes. If multiple imputation with auxiliary variables to handle the problem of missing data, the answer is no.

Missing data can also be handled using proxy variables. The use of proxies also has important implications for scientific research. It was showed that when parental education is used as a proxy for household income, it has a statistically significant effect in modelling teen pregnancy, but when household income was used in the equation the effect of parental education disappears.

This finding is very important for two reasons. First from a social policy perspective, the mechanisms and policies that can have an impact on income versus those that can have an effect on education are very different. Thus, knowing that the two variables have different effects on predicting the likelihood of an adolescent pregnancy depending on how one handles the problem of missing data is critical for conducting sociological research. Second, from a theoretical perspective, the use of proxies can have important implications because they might be measuring completely different constructs. For example, the health literature has shown that the effect of education on health is not the same as the effect of income on health (Mirowsky and Ross, 2003). Education taps human capital while income is restricted to financial resources (Sen, 1999). Therefore the effect of education versus that of income can potentially have very different effects on other models related to health outcomes.

This analysis has shown that missing data is indeed a critical component of scientific research, and that different techniques will often lead to different statistical and theoretical conclusions. The next logical question is, how are missing data to be handled when there are potential problems, even with the gold standard of multiple imputation. One of the best and most interesting responses to this question is: “The only good solution to missing data is not to have any” (Allison, 2001, p. 2). Because this is an unrealistic option, we propose that it is reasonable to ask researchers who are conducting analyses with missing data to report the results of both listwise deletion and multiple imputation. In addition, the researcher should try different methods of multiple imputation, i.e., with auxiliary variable and without them, to determine the level of consistency of the findings. Analyses with strong theories and consistent results across different methods of handling missing data should not be problematic. But when the findings are inconsistent, that is, they vary depending on how missing data is handled, and also when there is no strong theory, then the results should be rendered as inconclusive.

Finally, an important recommendation of our paper is that the effect of missing data on scientific research requires more scrutiny. The editors of peer reviewed journals should require the authors to report precisely the amount of data that is missing in their variables, as well as to specify and justify the method they used to handle missing data (Sterne et al., 2009). We specifically recommend that researchers should estimate their models with both listwise deletion and with multiple imputation and report if there are any differences that would lead to different theoretical or empirical conclusions. Research conducted with large amounts of missing data should be scrutinized with great deliberation and forethought, and the findings if inconsistent across method, should be interpreted with caution.

Acknowledgment

This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website

(<http://www.cpc.unc.edu/addhealth>). No direct support was received from grant P01-HD31921 for this analysis.

References

- Acock, A. (2005). Working with missing values. *Journal of Marriage and Family* 67, 1012-1028.
- Allison, P. D. (2001). *Missing Data*. Thousand Oaks, CA: Sage Publications.
- Bean, F. D., & Swicegood, G. (1985). *Mexican American Fertility Patterns*. Austin, TX: University of Texas Press.
- Enders, C. K. (2010). *Applied Missing Data Analysis*. New York: Guilford Press.
- Harris, K. M. (2008). *The National Longitudinal Study of Adolescent Health (Add Health), Waves I & II, 1994–1996; Wave III, 2001–2002 [machine-readable data file and documentation]*. Chapel Hill, NC: Carolina Population Center, University of North Carolina at Chapel Hill.
- Klepinger, D. H., Lundberg, S., & Plotnick, R. D. (1995). Adolescent fertility and the educational attainment of young women. *Family Planning Perspectives* 27, 23-28.
- Lee, K. J., & Carlin, J. B. (2010). Multiple imputation for missing data: Fully conditional specification versus multivariate normal imputation. *American Journal of Epidemiology*, 171, 624–632.
- Long, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage Publications.
- Long, J. S., & Freese, J. (2006). *Regression Models for Categorical Dependent Variables Using Stata. Second Edition*. College Station, TX: Stata Press.
- Mirowsky, J., & Ross, C. E. (2003). *Education, Social Status, and Health*. New York: A. de Gruyter.
- Poston, D. L., Jr. (2002). Son preference and fertility in China. *Journal of Biosocial Science*, 34, 333-347.
- Rosenbaum, J. E. (2006). Reborn a virgin: Adolescents' retracting of virginity pledges and sexual histories. *American Journal of Public Health*, 96, 1098-1103.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63, 581-590.

MISSING DATA AND THE STATISTICAL MODELING OF ADOLESCENT

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley and Sons.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. Boca Raton, FL: Chapman & Hall/CRC.

Sen, A. (1999). *Development As Freedom*. New York: Knopf.

StataCorp. (2011). *Stata Survey Data Reference Manual, Release 12*. College Station, TX: StataCorp.

Sterne, J. A., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., ... Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: Potential and pitfalls. *British Medical Journal*, 338(b2393). Retrieved from <http://www.bmj.com/content/338/bmj.b2393>

Treiman, D. J. (2009). *Quantitative Data Analysis: Doing Social Research to Test Ideas*. San Francisco: Jossey-Bass.

von Hippel, P. T. (2007). Regression with missing Ys: An improved strategy for analyzing multiple imputed data. *Sociological Methodology*, 37, 83-117.

Estimates and Forecasts of GARCH Model under Misspecified Probability Distributions: A Monte Carlo Simulation Approach

OlaOluwa S. Yaya

University of Ibadan
Ibadan, Nigeria

Olusanya E. Olubusoye

University of Ibadan
Ibadan, Nigeria

Oluwadare O. Ojo

Federal University of
Technology Akure
Akure, Nigeria

The effect of misspecification of correct sampling probability distribution of Generalized Autoregressive Conditionally Heteroscedastic (GARCH) processes is considered. The three assumed distributions are the normal, Student t , and generalized error distributions. The GARCH process is sampled using one of the distributions and the model is estimated based on the three distributions in each sample. Parameter estimates and forecast performance are used to judge the estimated model for performance. The AR-GARCH-GED performed better on the three assumed distributions; even, when Student t distribution is assumed, AR-GARCH-Student t does not perform as the best model.

Keywords: Generalized Error Distribution, forecasts, GARCH, misspecification, specification

Introduction

Since the introduction of Generalized Autoregressive Conditional Heteroscedastic (GARCH) model of Bollerslev (1986), thousands of articles have been published applying the model on financial series. The model captures volatility in the market, and its distributional specification makes it special among other nonlinear time series models. The GARCH process exists on the assumption of Normal, Student t , and Generalized Error Distributions (GED). The Normal distribution is the usual assumption in any time series estimation, but due to the fact that the distribution of GARCH process is leptokurtic, Normal distribution was found to be inappropriate

OlaOluwa Yaya and Olusanya Olubusoye are part of the academic staff in the Department of Statistics, University of Ibadan. Email them at os.yaya@ui.edu.ng. and oe.olubusoye@ui.edu.ng. O. O. Ojo is a lecturer in the Department of Statistics, Federal University of Technology, Akure. Email him at daruu208075@yahoo.com.

in capturing the tail behavior of the series. Bollerslev (1987) therefore proposed Student t distribution to capture the long tail behavior of the process. Nelson (1991) proposed the GED distribution.

Apart from the real applications of GARCH models on financial series, there is need to study the effect of misspecifying the GARCH distributional assumptions during estimation. Articles are very scarce along this line of thought. Wang (2002) affirms that spurious and inefficient inference is expected when pure GARCH models are misspecified. This as well may affect the Quasi Maximum Likelihood Estimates (QMLEs) of the misspecified model. The QMLE of pure GARCH(1,1) models indicates that the ARCH parameter is small, GARCH parameter is close to unity and the sum of both parameters approaches unity as the sampling frequency increases (Engle and Bollerslev, 1986; Bollerslev and Engle, 1993; Baillie, Bollerslev and Mikkelsen, 1996; Ding and Granger, 1996; Andersen and Bollerslev, 1997, and Engle and Patton, 2001.) This fact is reflected in the Integrated GARCH (IGARCH) of Engle and Bollerslev (1986). A more recent paper by Jensen and Lange (2010) shows that in a GARCH (1,1) model, the estimates of $\hat{\alpha}_1$ and $\hat{\beta}_1$ tend to zero and unity respectively as the sampling frequency increased, which is an IGARCH effect. This IGARCH effect is known for pure-GARCH processes. In a linear AR-GARCH or nonlinear AR-GARCH processes, IGARCH effect is not plausible. The present work considers AR-GARCH process, and therefore IGARCH effect may not be expected.

As tail distribution of the GARCH model is captured using the three distributions, and parameters estimated adjust accordingly, forecasts performances of the model are affected. Extensive Monte Carlo simulation was performed on the GARCH model using the three distributions.

The GARCH (1,1) model

The GARCH (1,1) model proposed in Bollerslev (1986) is

$$\sigma_t^2 = w + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (1)$$

where ε_t are the returns series of the financial asset; σ_t is the volatility at time t and z_t gives the assumed distribution. The parameters, α_1 and β_1 are conditioned as $w > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$, and $\alpha_1 + \beta_1 < 1$ in order to ensure stationarity of the whole process (Bollerslev, 1986). This condition is establish by defining

$k_t = \varepsilon_t^2 - \sigma_t^2 = (z_t^2 - 1)\sigma_t^2$ where $z_t \approx N(0,1)$. Using this in (1) results in the Autoregressive Moving Average (ARMA) representation

$$\varepsilon_t^2 = w + (\alpha_1 + \beta_1)\varepsilon_{t-1}^2 - \beta_1 k_{t-1} + k_t \quad (2)$$

where k_t is serially uncorrelated with mean zero. Stationarity of the process is then ensured when the roots of $1 - \alpha(1) - \beta(1) = 1 - (\alpha_1 + \beta_1) = 0$ lie outside the unit circle and this is not conditioned on time t as it is measured directly from the parameters of the model. Hence it is expected that

$$(\alpha_1 + \beta_1) < 1 \quad (3)$$

for existence of covariance stationary process. For the stationary process, the finite unconditional variance of ε_t is given by

$$\sigma^2 = \frac{w}{1 - \alpha_1 - \beta_1} \quad (4)$$

Kurtosis of GARCH (1,1) model

For any GARCH (p,q) process, $E(z_t) = 0$ and $Var(z_t) = 1$. $E(z_t^3)$ is the skewness and $E(z_t^4)$ gives the measure of skewness. Because the emphasis is on tail behaviour of GARCH residuals, the expression for the unconditional kurtosis is next derived.

Assuming that $E(\sigma_t^2)$ and $E(\sigma_t^4)$ exist, then it suffices to write

$$E(z_t^4) = k_z + 3;$$

because $E(\varepsilon_t) = 0$;

$$E(\varepsilon_t^4) = E(\sigma_t^4)E(z_t^4) \text{ because } \varepsilon_t = \sigma_t z_t \text{ and } \varepsilon_t \text{ and } z_t \text{ are independent}$$

ESTIMATES OF GARCH MODEL

Then, squaring GARCH (1,1) model,

$$\sigma_t^2 = w + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

gives

$$\sigma_t^2 = w^2 + \alpha_1^2 \varepsilon_{t-1}^4 + \beta_1^2 \sigma_{t-1}^4 + 2w\alpha_1 \varepsilon_{t-1}^2 + 2w\beta_1 \sigma_{t-1}^2 + 2\alpha_1 \beta_1 \sigma_{t-1}^2 \varepsilon_{t-1}^2.$$

Taking expectation of the resulting expansion, as well as applying the properties outlined above

$$E(\sigma_t^2) = \frac{w^2(1 + \alpha_1 + \beta_1)}{[1 - (\alpha_1 + \beta_1)][1 - \alpha_1^2(k_z + 2) - (\alpha_1 + \beta_1)^2]}$$

Using the relation $E(\varepsilon_t^4) = (k_z + 3)E(\sigma_t^4)$ where k_z is the excess kurtosis of z_t , then

$$E(\varepsilon_t^4) = \frac{w^2(1 + \alpha_1 + \beta_1)(k_z + 3)}{[1 - (\alpha_1 + \beta_1)][1 - \alpha_1^2(k_z + 2) - (\alpha_1 + \beta_1)^2]}$$

Using the formula $k_\varepsilon = \frac{E(\varepsilon_t^4)}{[E(\varepsilon_t^2)]^2} - 3$ for excess kurtosis and with the fact that

$$E(\varepsilon_t^2) = \frac{w}{1 - \alpha_1 - \beta_1} \text{ from the properties above,}$$

$$\begin{aligned} k_\varepsilon &= \frac{w^2(1 + \alpha_1 + \beta_1)(k_z + 3)[1 - (\alpha_1 + \beta_1)]^2}{w^2[1 - (\alpha_1 + \beta_1)][1 - \alpha_1^2(k_z + 2) - (\alpha_1 + \beta_1)^2]} - 3 \\ &= \frac{(k_z + 3)[1 - (\alpha_1 + \beta_1)^2]}{[1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2 - k_z(\alpha_1^2)]} - 3 \end{aligned} \quad (5)$$

with normally distributed innovations z_t , $k_z = 0$ and

$$k_\varepsilon = \frac{6\alpha_1^2}{1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2} \quad (6)$$

with non-normally distributed innovations z_t , as in Student t and GED,

$$\text{Var}(z_t) = E(z_t^2) = 1 \text{ and } k_z = \frac{E(z_t^4)}{[E(z_t^2)]^2} = E(z_t^4) \neq 0, \text{ then}$$

$$k_\varepsilon = \frac{k_z - k_z(\alpha_1 + \beta_1)^2 + 6\alpha_1^2 + 2k_z\alpha_1^2}{1 - 2\alpha_1^2 - (\alpha_1 + \beta_1)^2 - k_z(\alpha_1^2)} \quad (7)$$

In these two cases, it is observed that α_1^2 is important in determining the tail behavior of ε_t , because once $\alpha_1^2 = 0$, $k_\varepsilon = 0$. Hence, $k_\varepsilon = k_z$ for the non-normally distributed case and it implies the similarity of the tail behaviors of both ε_t and z_t

Distributional Assumptions and Estimation

For GARCH models, the unconditional distributions are always non normal, and this gives fatter tails than the normal distribution. In practice, z_t is assumed to follow the normal distribution or non-normal distributions. These non-normal distributions have been proved to perform well in modeling the fatter tails (leptokurticity) observed in GARCH residuals. The non-normal distributions are the Student t distribution proposed in Bollerslev (1987) and Generalized Error Distribution (GED) by Nelson (1991)

The standardized Normal distribution is

$$f(z_t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z_t^2\right), -\infty < z_t < \infty \quad (8)$$

with the log likelihood function

$$L(z_t) = -\frac{1}{2} \left[N \log(2\pi) + \sum_{t=1}^N z_t^2 \right] \quad (9)$$

where N is the sample size. The standardized Student t distribution proposed in Bollerslev (1987) is given as

$$f(z_t, \nu) = \Gamma((\nu+1)/2) [\Gamma(\nu/2)]^{-1} [\pi(\nu)]^{-1/2} \left(1 + \frac{z_t^2}{\nu}\right)^{-(\nu+1)/2}, \quad -\infty < z_t < \infty \quad (10)$$

This distribution is symmetric around zero as it is observed in its specification with $\nu > 2$. At $\nu = 1$, the Student t reduces to Cauchy distribution. At $2 < \nu \leq 4$, its conditional kurtosis is less than 3, which means that the resulting tail effect is normal. For $\nu > 4$, the kurtosis becomes $3(\nu-1)(\nu-4)^{-1}$, which is greater than 3, hence the tail effect becomes non-normal distribution. As $\nu \rightarrow \infty$, the distribution converges to normal distribution. The log likelihood function of Student t distribution is then simplified as

$$L(z_t, \nu) = -\frac{1}{2} \left\{ N \log \left(\frac{\pi(\nu-2)\Gamma(\nu/2)^2}{\Gamma((\nu+1)/2)^2} \right) + (\nu+1) \sum_{t=1}^N \log \left[1 + \frac{z_t^2}{(\nu-1)} \right] \right\} \quad (11)$$

The standardized GED proposed in Nelson (1991) is given as

$$f(z_t, \nu) = 2^{-1} \nu \Gamma\left(\frac{3}{\nu}\right)^{1/2} \left[\Gamma\left(\frac{1}{\nu}\right)^{3/2} \right]^{-1} \exp \left[-|z_t|^\nu \left\{ \Gamma\left(\frac{3}{\nu}\right) \left[\Gamma\left(\frac{1}{\nu}\right) \right]^{-1} \right\}^{\nu/2} \right] \quad (12)$$

where $-\infty < z_t < \infty$ and $\nu > 0$. The GED reduces to the standard normal distribution at $\nu = 4$. At $0 < \nu < 2$, the distribution has thicker tail than the normal distribution, for example, at $\nu = 1$ the distribution becomes a double exponential (Laplace) distribution. At $\nu > 2$, the distribution of z_t has thinner tails than the normal distribution, for example, as ν tends to infinity, z_t reduces to a uniform distribution on the interval $(-\sqrt{3}, \sqrt{3})$. The log likelihood of this distribution is then expanded as

$$f(z_t, \nu) = 2^{-1} \nu \Gamma\left(\frac{3}{\nu}\right)^{1/2} \left[\Gamma\left(\frac{1}{\nu}\right)^{3/2} \right]^{-1} \exp \left[-|z_t|^\nu \left\{ \Gamma\left(\frac{3}{\nu}\right) \left[\Gamma\left(\frac{1}{\nu}\right) \right]^{-1} \right\}^{\nu/2} \right] \quad (13)$$

These likelihood functions are then estimated using the numerical derivatives based on the fact that GARCH models lack closed form estimation. Berndt, Hall, Hall and Hausman (BHHH) algorithm of Berndt, et al (1974) is then used. This algorithm is termed Gauss-Newton in general Nonlinear Least Squares (NLS) and BHHH in MLE estimation. Unlike some other derivatives, it uses only first derivatives of the likelihood function and computes a set of parameter values as

$$\psi^{(i+1)} = \psi^{(i)} - \left(\sum_{t=1}^N \frac{\partial L(z_{t,\cdot})^{(i)}}{\partial \psi} \cdot \frac{\partial L(z_{t,\cdot})^{(i)}}{\partial \psi'} \right)^{-1} \frac{\partial L(z_{t,\cdot})^{(i)}}{\partial \psi} \quad (14)$$

where $L(z_{t,\cdot})$ is the likelihood function. The initial parameter set is given as $\psi^{(0)}$ and the parameter set which maximize the likelihood function is denoted as $\psi^{(i+1)}$. The estimation of GARCH (1, 1) model with Student t distribution and GED follow the usual Quasi Maximum Likelihood Estimation (QMLE) because normality assumption is violated in these cases.

Misspecification of distribution of GARCH model could lead to stationarity and explosion of the series in some points. Though standard errors will be consistent; the QML estimators $\tilde{\psi}^{(i+1)}$ are generally closed to the exact ML estimator $\hat{\psi}^{(i+1)}$ for symmetric GARCH distribution. For non-symmetric conditional distributions, both the asymptotic and finite sample loss in efficiency are quite large and parametric estimation approach are not applicable in this regard (Mills and Markellos, 2008).

Forecasts Evaluation

Forecast evaluation criteria considered are the Root Mean Squares Forecast Error (RMSFE), Mean Absolute Error (MAE), Mean Absolute Percentage Forecast Error (MAPFE) and Theil Inequality of Theil (1961;1966). The MSFE is defined as

$$MSFE = \frac{1}{m} \sum_{t=1}^m (\hat{\sigma}_t^2 - \sigma_t^2)^2 \quad (15)$$

where $\hat{\sigma}_t^2$ is the predicted in-sample conditional variances, and this depends on the scale of the variance series, σ_t^2 . The square root of MSFE is the RMSFE

ESTIMATES OF GARCH MODEL

$$RMSFE = \sqrt{\frac{1}{m} \sum_{t=1}^m (\hat{\sigma}_t^2 - \sigma_t^2)^2} \quad (16)$$

The MAFE and MAPFE are obtained by taking the absolute differences of the predicted conditional volatilities and the observed volatilities as

$$MAFE = \frac{1}{m} \sum_{t=1}^m |\hat{\sigma}_t^2 - \sigma_t^2| \quad (17)$$

$$MAPFE = 100 \sum_{t=1}^m \left| \frac{\hat{\sigma}_t^2 - \sigma_t^2}{\sigma_t^2} \right| \quad (18)$$

The Theil inequality is given as

$$TI = \frac{\sqrt{\frac{1}{m} \sum_{t=1}^m (\hat{\sigma}_t^2 - \sigma_t^2)^2}}{\sqrt{\frac{1}{m} \sum_{t=1}^m \hat{\sigma}_t^2} + \sqrt{\frac{1}{m} \sum_{t=1}^m \sigma_t^2}} \quad (19)$$

The inequality coefficient is time invariant and always lies between 0 and unity. The smaller these forecast evaluation criteria, the better the candidate model represent well the data.

Monte Carlo Simulations

The Monte Carlo experiment is set up using the AR(1)–GARCH(1,1) DGP

$$y_t = 0.15 + 0.5y_{t-1} + \varepsilon_t, \quad \sigma_t^2 = 0.02 + 0.25\varepsilon_{t-1}^2 + 0.60\sigma_{t-1}^2, \quad (20)$$

with the error distribution $\varepsilon_t = \sigma_t z_t$ where z_t is assumed to follow Normal, Student t and GED distributions. The parameters of the AR(1) and GARCH(1,1) models are set within the stationary region in order to avoid problems data explosion. The sample sizes N are varied as 2000, 4000 and 6000 with in-sample forecasts generated as 25% of the data length. The results are then presented as Scenarios 1

to 3 in Tables 1–6 below. Each Scenario gives results for parameter estimation, volatility, excess kurtosis and forecasts evaluations criteria.

Scenario 1: When the true Distribution is Normal

Tables 1 and 2 present the results when the GARCH processes are simulated based on Normal distribution assumption, and these processes are used to estimate the GARCH process based on Student t , GED and the same Normal distribution. The results in Table 1 show that the AR-GARCH parameter estimates, measures of volatility and kurtosis are not consistent with sample sizes. Both the AR and GARCH parameter estimates computed for Student t distribution have larger biases in compared with that of Normal and GED distributions, even though excess kurtosis of the AR-GARCH-Student t model is the smallest. Volatility of the AR-GARCH-Student t model is also observed to be higher than that of the Normal and GED distributions. The excess kurtosis of the AR-GARCH-Normal model was expected to be the smallest because the series is sampled from Normal distribution but this was not the case.

Looking at the results of the in-sample forecasts realized from the AR-GARCH models as given in Table 2, the AR-GARCH-Normal and AR-GARCH-GED model perform better than AR-GARCH-Student t model on forecasts as given by the minimum values of the RMSPE and Theil inequality coefficients. The AR-GARCH-GED is expected to realize better forecasts than AR-GARCH-Normal model.

Table 1. Model Parameter, Volatility and Kurtosis when GARCH processes are simulated based on Normal distribution assumption

Assumed Distribution	Sample	$\hat{\phi}_0$ (0.1500)	$\hat{\phi}_1$ (0.5000)	\hat{w} (0.0200)	$\hat{\alpha}_1$ (0.2500)	$\hat{\beta}_1$ (0.6000)	Persistence (0.8500)	Volatility	Exc. Kurtosis
Normal	2000	0.1480	0.4839	0.0169	0.2110	0.6596	0.8706	0.1306	1.6427
	4000	0.1518	0.4724	0.0173	0.2049	0.6590	0.8639	0.1271	1.8430
	6000	0.1475	0.4750	0.0180	0.2052	0.6503	0.8555	0.1246	1.3275
Student t	2000	0.1462	0.4868	0.0820	0.1500	0.6000	0.7500	0.3280	0.2895
	4000	0.1471	0.4794	0.0794	0.1500	0.6000	0.7500	0.3176	1.8408
	6000	0.1576	0.4916	0.0849	0.1500	0.6000	0.7500	0.3396	0.4501
GED	2000	0.1499	0.4811	0.0160	0.2125	0.6831	0.8956	0.1533	2.3761
	4000	0.1547	0.4662	0.0172	0.2106	0.6743	0.8849	0.1494	2.1683
	6000	0.1495	0.4723	0.0185	0.2118	0.6612	0.8730	0.1457	1.7707

ESTIMATES OF GARCH MODEL

Table 2. Forecast evaluation estimates when GARCH processes are simulated based on Normal distribution assumption

Assumed Distribution	Sample	RMSPE	MAFE	MAPFE	Theil
Normal	2000	0.0003	0.0112	48.0967	0.0289
	4000	0.0002	0.0104	44.8820	0.0257
	6000	0.0001	0.0101	49.5855	0.0287
Student t	2000	0.0011	0.0464	500.3362	0.1270
	4000	0.0007	0.0437	529.5600	0.1242
	6000	0.0008	0.0602	503.3612	0.1363
GED	2000	0.0003	0.0110	50.1820	0.0276
	4000	0.0002	0.0102	47.7504	0.0246
	6000	0.0001	0.0099	53.1115	0.0276

Scenario 2: When the true Distribution is Student t

Tables 3 and 4 present the results when the true GARCH distribution follows Student t . Here, the distinctions in the GARCH estimates can only be made using the persistence and unconditional volatility measures. The AR-GARCH-Student t model still presents smallest persistence and highest volatility. The excess kurtosis of the AR-GARCH-Student t model is the smallest followed by that of AR-GARCH-Normal model.

Table 3. Model Parameter, Volatility and Kurtosis when the true GARCH distribution follows Student t

Assumed Distribution	Sample	$\hat{\phi}_0$ (0.1500)	$\hat{\phi}_1$ (0.5000)	\hat{w} (0.0200)	$\hat{\alpha}_1$ (0.2500)	$\hat{\beta}_1$ (0.6000)	Persistence (0.8500)	Volatility	Exc. Kurtosis
Normal	2000	0.1408	0.5096	0.0232	0.2295	0.6105	0.8400	0.1450	2.5966
	4000	0.1478	0.4830	0.0237	0.2576	0.5762	0.8338	0.1426	4.0384
	6000	0.1497	0.4878	0.0219	0.2483	0.6032	0.8515	0.1475	3.7926
Student t	2000	0.1510	0.5165	0.0938	0.1500	0.6000	0.7500	0.3752	0.7336
	4000	0.1472	0.4829	0.0911	0.1500	0.6000	0.7500	0.3644	0.8526
	6000	0.1408	0.5114	0.0961	0.1500	0.6000	0.7500	0.3844	1.1013
GED	2000	0.1416	0.4967	0.0261	0.2476	0.5979	0.8455	0.1689	3.5293
	4000	0.1473	0.4792	0.0243	0.2620	0.5937	0.8557	0.1684	5.7359
	6000	0.1491	0.4890	0.0230	0.2585	0.6093	0.8678	0.1740	5.9655

Table 4. Forecast evaluation estimates when the true GARCH distribution follows Student t

Assumed Distribution	Sample	RMSPE	MAFE	MAPFE	Theil
Normal	2000	0.0004	0.0145	63.7385	0.0332
	4000	0.0002	0.0131	62.2680	0.0324
	6000	0.0003	0.0247	63.1032	0.0392
Student t	2000	0.0015	0.0627	635.1114	0.1552
	4000	0.0010	0.0574	611.0665	0.1471
	6000	0.0009	0.0678	645.0181	0.1582
GED	2000	0.0004	0.0156	76.3588	0.0355
	4000	0.0002	0.0132	68.0163	0.0321
	6000	0.0004	0.0253	70.6578	0.0396

In terms of forecasts, the AR-GARCH-Student t model is the worst, even though the DGP is realized from the same probability distribution. The forecast performances of AR-GARCH-Normal and AR-GARCH-GED seem not different from each other as indicated by the forecast evaluation estimates.

Scenario 3: When the true Distribution is GED

Table 5 and 6 present the results when the true GARCH distribution is GED. In Table 5, in the AR(1) estimates, the estimates for the constant $\hat{\phi}_0$ are all consistent with sample sizes when the three probability distributions are assumed. The autoregressive parameters $\hat{\phi}_1$ are not consistent with sample sizes. The GARCH parameter estimates computed for Student t distribution are the same to that of Table 1 and 3 while the AR(1) parameter are different in the two results. The Student t distribution assumption of GARCH model still presents model estimates with highest volatility but with lowest persistence of this volatility. Misspecifying GED for Student t distribution here also caused the excess kurtosis to be negative in AR-GARCH-Student t model and this is a very spurious case.

The forecast evaluation results of the model estimates follow in In Table 6. Starting with the AR(1)–GARCH(1,1)–Student t model, the model is the worst in terms of forecasts because it presents the highest RMSPE, MAPE, MAPFE and Theil inequality coefficient. The best model is AR(1)–GARCH(1,1)–GED model, and this is expected because the DGP assumed GED initially. The performance of

ESTIMATES OF GARCH MODEL

AR(1)–GARCH(1,1)–Normal in terms of forecast is very close to that of AR(1)–GARCH(1,1)–GED model.

Table 5. Model Parameter, Volatility and Kurtosis when the true GARCH distribution is GED

Assumed Distribution	Sample	$\hat{\phi}_0$ (0.1500)	$\hat{\phi}_1$ (0.5000)	\hat{w} (0.0200)	$\hat{\alpha}_1$ (0.2500)	$\hat{\beta}_1$ (0.6000)	Persistence (0.8500)	Volatility	Exc. Kurtosis
Normal	2000	0.1409	0.5120	0.0219	0.2260	0.6112	0.8372	0.1345	0.1984
	4000	0.1478	0.4848	0.0236	0.2576	0.5639	0.8215	0.1322	0.4768
	6000	0.1497	0.4880	0.0218	0.2465	0.5932	0.8397	0.1360	0.3859
Student <i>t</i>	2000	0.1373	0.5192	0.0877	0.1500	0.6000	0.7500	0.3508	-0.3883
	4000	0.1469	0.4863	0.0854	0.1500	0.6000	0.7500	0.3416	-0.3554
	6000	0.1500	0.4860	0.0875	0.1500	0.6000	0.7500	0.3500	-0.4099
GED	2000	0.1424	0.5024	0.0225	0.2284	0.6031	0.8315	0.1335	0.1874
	4000	0.1481	0.4834	0.0228	0.2493	0.5763	0.8256	0.1307	0.4139
	6000	0.1498	0.4900	0.0214	0.2430	0.5973	0.8403	0.1340	0.3518

Table 6. Forecast evaluation estimates when the true GARCH distribution is GED

Assumed Distribution	Sample	RMSPE	MAFE	MAPFE	Theil
Normal	2000	0.0002	0.0099	48.9918	0.0280
	4000	0.0002	0.0100	50.8079	0.0292
	6000	0.0002	0.0142	50.9182	0.0325
Student <i>t</i>	2000	0.0013	0.0529	521.4746	0.1396
	4000	0.0008	0.0496	499.5437	0.1329
	6000	0.0008	0.0547	494.8839	0.1379
GED	2000	0.0002	0.0100	49.5704	0.0284
	4000	0.0002	0.0097	48.9723	0.0284
	6000	0.0002	0.0140	49.9722	0.0322

Conclusion

The misspecification of GARCH probability distribution functions were considered. These are the Normal, Student t and Generalized Error Distributions (GED). The estimation convergence time varied based on the distribution and the set sample sizes. When a Normal distribution was assumed, the AR–GARCH–GED seemed to perform marginally better than AR–GARCH–Normal model in terms of forecasts as revealed in the estimates of the Theil inequality. Though, the AR–GARCH–Normal was the best model here in terms of parameter estimates, and this was expected because the DGP assumed Normal distribution initially. With the assumption of Student t distribution in the DGP, the forecast performance of the models computed with Normal distribution and GED reduced and these still presented better models than the corresponding AR–GARCH–Student t model. Similar results were obtained when the DGP assumed GED.

It was also observed that all the results obtained, particularly the parameter estimates were not consistent with sample sizes. These are expected because volatility came into play. In empirical modeling research like this, interest should either lie in the behavior of the volatility—assuming a probability distribution which will give us the best volatility measurement—or in the forecasts. The best GARCH model may not actually produce the best forecast estimates and probability distributions have effect on the tail distribution of the innovations. This work can be replicated using higher order of the model, and in that case, more sophisticated software is recommended for the simulation in order to avoid convergence problems.

References

- Andersen, T. G., & Bollerslev, T. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, 4, 115-158.
- Baillie, R.T., Bollerslev, T., & Mikkelsen, H. O. (1996). Fractional integrated generalized autoregressive conditional heteroscedasticity, *Journal of Econometrics*, 74, 3-30.
- Berndt, E., Hall, B., Hall, R., & Hausman, J. (1974). Estimation and inference in nonlinear structural models. *Annals of Economic and Social Measurement*, 3, 653-665.
- Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics*, 31, 307–327.

ESTIMATES OF GARCH MODEL

- Bollerslev, T. (1987). A conditional heteroscedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics*, 69, 542-547.
- Bollerslev, T., & Engle, R. F. (1993). Common persistence in conditional variances. *Econometrica*, 61, 167-186.
- Ding, Z., & Granger, C. W. J. (1996), Modelling volatility persistence of speculative returns: a new approach. *Journal of Econometrics*, 73, 185-215.
- Engle, R. F., & Bollerslev, T. (1986). Modeling the persistence of conditional variances. *Econometric Reviews*, 5, 1-50.
- Engle, R. F., & Patton, A. J. (2001). What good is a volatility model? *Quantitative Finance*, 1, 237-245.
- Jensen, A. T., & Lange, T. (2010). On convergence of the QMLE for misspecified GARCH models. *Journal of Time Series Econometrics*, 2(1), 1-31.
- Mills, T. C., & Markellos, R. (2008). *The Econometric Modelling of Financial Time Series* (3rd ed.). Cambridge, UK: Cambridge University Press.
- Nelson, D. B. (1991). Conditional heteroscedasticity in asset returns: A new approach. *Econometrica*, 59, 347-370.
- Theil, H. (1961). *Economic forecasts and policy*. Amsterdam: North-Holland.
- Theil, H. (1966). *Applied Economic forecasting*. Amsterdam: North-Holland
- Wang, Y. (2002), Asymptotic nonequivalence of GARCH models and diffusions. *Annals of Statistics*, 30(3), 754-783.

Emerging Scholars: Bayesian Inference for Volatility of Stock Prices

Juliet Gratia D'Cunha

Mangalore University
Mangalagangothri, Karnataka, India

K. Aruna Rao

Mangalore University
Mangalagangothri, Karnataka, India

Lognormal distribution is widely used in the analysis of failure time data and stock prices. Maximum likelihood and Bayes estimator of the coefficient of variation of lognormal distribution along with confidence/credible intervals are developed. The utility of Bayes procedure is illustrated by analyzing prices of selected stocks.

Keywords: Bayesian inference, volatility, stock prices, coefficient of variation, lognormal distribution

Introduction

The study on coefficient of variation (CV) of the normal distribution dates back to McKay (1932); since then various articles have appeared concerning improved estimation of CV of a normal distribution and tests for equality of CV's of two or more normal distributions. Some of the recent references regarding the estimation of CV of the normal distribution are Ahmed (1995), Breunig (2001), Liu, et al. (2006), Mohmoudvand & Hassani (2009) and Panichkitkosolkul (2009). The papers dealing with tests for equality of CV's of independent normal distributions are Bennett (1976), Doornabos & Dijkstra (1983), Shafer & Sullivan (1986), Gupta & Ma (1996), Nairy & Rao (2003) and Verril & Johnson (2007). In addition to these papers, the papers on CV relating to finance and economics are Brief & Owen (1969), Jobson & Korkie (1981), De, et al. (1996) and Memmel (2003). These papers are developed on the assumption of normality of the observations.

Juliet Gratia D'Cunha is a Research Scholar in the Department of Statistics. Email her at gratiajuliet@gmail.com. Dr. Rao is a Professor in the Department of Statistics. Email him at arunaraomu@gmail.com.

BAYESIAN INFERENCE FOR VOLATILITY OF STOCK PRICES

Generally stock prices do not follow normal distribution and the data is analyzed using logarithm of prices. This amounts to the assumption that stock price is lognormally distributed. CV is not invariant under distributional transformation, and thus estimators are to be derived for the CV of the lognormal distribution.

Maximum likelihood estimator (M.L.E) and confidence interval for the CV of the lognormal distribution are derived, as well as the Bayes estimator of CV of the lognormal distribution using a) Right invariant prior b) Left invariant Jeffrey's prior.

Bayesian inference has several advantages over the likelihood based inference (Ghosh, et al., 2006; Berger, 1985). Simulation study carried out in this paper suggests that Bayesian credible intervals have smaller average length compared to the confidence interval obtained by M.L.E. Financial analysts are generally not well exposed to Bayesian analysis and this paper introduces this idea by analyzing the stock prices of 3 Indian stocks.

The maximum likelihood estimator and Bayes estimator of the CV of the lognormal distribution and the associated confidence/credible intervals are initially derived. A simulation study is conducted to compare the coverage probability and average length of the confidence/credible intervals. The procedures developed in this paper are illustrated by analyzing stock prices of 3 scripts belonging to large cap sector of the Indian stock market. For this purpose daily data from August 19 to November 6 for the year 2013 is used. By using part of the data as training set and remaining data as the validation set, the paper demonstrates that Bayesian inference can be used to predict stock market volatility.

Bayes Estimator of CV of the Lognormal Distribution

Let x_1, x_2, \dots, x_n be a random sample from lognormal distribution with density

$$f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\log x - \mu)^2}{2\sigma^2}}, x > 0, -\infty < \mu < \infty, \sigma > 0 \quad (1)$$

Denoting $\log X_i$ as Z_i , the minimal sufficient statistic for μ and σ^2 are

$$\bar{Z} = \frac{\sum_{i=1}^n Z_i}{n} \quad (2)$$

and

$$S_z^2 = \frac{\sum_{i=1}^n (Z_i - \bar{Z})^2}{n} \quad (3)$$

Therefore the maximum likelihood estimator of μ and σ^2 are \bar{Z} and $\left(\frac{n-1}{n}\right)S_z^2$. The mean, variance and coefficient of variation of the lognormal distribution are

$$E(x) = e^{\mu + \frac{1}{2}\sigma^2} \quad (4)$$

$$V(x) = \left(e^{2\mu + \sigma^2}\right)\left(e^{\sigma^2} - 1\right) \quad (5)$$

$$CV = \left(e^{\sigma^2} - 1\right)^{\frac{1}{2}} \quad (6)$$

respectively. Using the invariance property of maximum likelihood estimators, the maximum likelihood estimator of the CV of lognormal distribution is given by

$$\left(e^{S_z^2} - 1\right)^{\frac{1}{2}}, \text{ (Calculations herein used } \frac{n-1}{n} \approx 1) \quad (7)$$

The Bayes estimator of the CV of the lognormal distribution depends on the specification of the prior distribution for μ and σ^2 . In objective Bayesian analysis, the commonly used priors are the following

- *Right invariant prior:* For the location scale family with location parameter μ and scale parameter σ , the right invariant prior is $\pi(\mu, \sigma) = 1/\sigma$.
- *Jeffrey's prior:* Jeffrey's prior for μ and σ is given by $\pi(\mu, \sigma) = 1/\sigma^2$. Jeffrey's prior is left invariant but not right invariant.

Because the lognormal distribution belongs to log location scale family, the above priors were used in this study. Although right invariant prior is recommended (Ghosh, et al., 2006; Berger, 1985), the use of Jeffrey's prior aids in studying the

BAYESIAN INFERENCE FOR VOLATILITY OF STOCK PRICES

Bayesian robustness with respect to specification of the prior distribution. Because the distribution of \bar{Z} and S_z^2 are independent, denoting $\eta = 1/\sigma^2$, after some simplification the posterior density of η is obtained as $\text{Gamma}\left(\frac{n+2}{2}, \frac{1}{2}(n-1)S_z^2\right)$ is obtained under right invariant prior and $\text{Gamma}\left(\frac{n+3}{2}, \frac{1}{2}(n-1)S_z^2\right)$ under Jeffrey's prior.

Under squared error loss function, the Bayes estimator of CV is

$$E\left(\left(e^{\sigma^2} - 1\right)^{\frac{1}{2}}\right) = E\left(\left(e^{\frac{1}{\eta}} - 1\right)^{\frac{1}{2}}\right) \quad (8)$$

where the expectation is taken with respect to the posterior density of $\pi(\eta|z)$. This expectation must be evaluated numerically, thus the importance sampling approach was used to evaluate the integral. In this approach observations are generated from the posterior density and the numerical value of the expectation is given by

$$E\left(e^{\sigma^2} - 1\right)^{\frac{1}{2}} = \frac{1}{M} \sum_{i=1}^M \left(e^{\sigma_i^2} - 1\right)^{\frac{1}{2}} \quad (9)$$

where $\sigma_i^2, i=1$ to M refers to the value of $1/\eta_i$ generated from the posterior density and M denotes the number of sample values generated. 10,000 observations are generated from the posterior density and using this, the Bayes estimator and equi-tailed credible intervals are obtained. For the likelihood based confidence interval, the equi-tailed confidence interval for $\eta = 1/\sigma^2$ is constructed using the Chi-square distribution for $\frac{(n-1)}{\sigma^2} S_z^2$. This confidence interval is then inverted to give a confidence interval for CV of the lognormal distribution. The confidence interval based on maximum likelihood estimator is given by

$$\left[\left(e^{\sigma_{\hat{L}}^2} - 1\right)^{\frac{1}{2}}, \left(e^{\sigma_{\hat{U}}^2} - 1\right)^{\frac{1}{2}} \right] \quad (10)$$

where

$$\sigma^2_L = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\chi^2_{\frac{\alpha}{2}}(n-1)} \quad (11)$$

and

$$\sigma^2_U = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\chi^2_{(1-\frac{\alpha}{2})}(n-1)} \quad (12)$$

Finite Sample Comparison of Credible and Confidence Intervals

The advantage of Bayes inference over likelihood-based inference is that it gives straightforward interpretation of the credible interval. Nevertheless, the superiority of the Bayes inference follows by comparing the coverage probability and length of the credible interval compared to the confidence interval based on maximum likelihood estimator. For this purpose a simulation study is conducted. For a random sample of size n ($n = 10, 20, 40, 60, 80, 100, 150, 200$, observations are generated from lognormal distribution or equivalently from normal distribution) with parameter μ and σ^2 . The value of μ and σ^2 are adjusted to yield a CV of 0.1, 0.3, 0.5, 0.7, 1, 1.5, 2, 2.5. The value of μ is fixed at 3. For the sample size and the value of CV, maximum likelihood estimator and the associated confidence intervals are computed using the expressions given in the previous section. For this sample size and value of CV, Bayes estimator, equi-tailed and HPD credible intervals are obtained using 10,000 simulated values of η , and thereby $(e^{\frac{1}{\eta}} - 1)^{\frac{1}{2}}$ from the posterior gamma density of η . This constitutes a single run in the simulation experiment. In each run the length of the confidence/credible interval is recorded. In addition, it is also recorded that whether the true value lies inside the confidence/credible interval. To estimate the coverage probability and average length of the confidence interval, the simulation experiment is repeated using 1000 runs. The coverage probability refers to the proportion of times the true value lies inside the interval. The credible/confidence level is fixed at 0.95. Tables 1 and 2 summarize the results of the simulation study.

BAYESIAN INFERENCE FOR VOLATILITY OF STOCK PRICES

Table 1. Coverage probability of the credible and confidence interval for the CV across sample sizes for 8 combinations of specified values of CV

Sample Size	Bayes Procedure (Equi-tailed)				Maximum Likelihood (Equi-tailed)	
	# of times Coverage probability Is maintained		Average length		# of times Coverage probability Is maintained	Average length
	Right invariant prior	Jeffrey's prior	Right invariant prior	Jeffrey's prior		
10	0	0	*	*	8	19.0641
20	0	0	*	*	8	2.4722
40	0	0	*	*	8	1.1390
60	1	0	1.4965	*	8	0.8264
80	4	0	0.1812	*	8	0.6888
100	8	0	0.5513	*	8	0.5976
150	8	7	0.4472	0.5010	8	0.4715
200	7	5	0.4363	0.4342	7	0.4477
Overall	28	12	0.6225	0.4676	63	3.2134

* Whenever coverage probability is not maintained average length has not been calculated

It may be said that the coverage probability is maintained if the estimated coverage probability lies between 0.940 to 0.960. That is $(1-\alpha) \pm 0.01$. From the table it is clear that the confidence interval based on maximum likelihood estimator maintains coverage probability for all sample sizes. On the other hand the equi-tailed credible interval maintains coverage probability when the sample size is greater than or equal to 100. However the average length of the credible interval is much shorter compared to the confidence interval. For example when $n = 150$ using right invariant prior, the average length of the credible interval is 0.4472 and using Jeffrey's prior it is 0.5010 while for the confidence interval it is 0.4715. The average length of the interval is computed using those intervals for which the coverage probability is maintained. The length of the confidence interval for Jeffrey's prior is marginally higher than right invariant prior. Table 2 presents the coverage probability and length of the HPD credible interval.

Table 2 shows that HPD credible interval maintains coverage probability when the sample size is greater than or equal to 40. The average length of the HPD credible intervals for both right and left invariant priors is marginally larger than the equi-tailed credible intervals. Theoretically the length of the HPD credible interval should be shorter than equi-tailed credible interval. To explore the reason for this phenomenon the posterior density for sample size $n = 60$ and 100 were plotted and the histogram and frequency curve of the simulated distribution of $\left(e^{\frac{1}{\eta}} - 1\right)^{\frac{1}{2}}$ was also plotted.

Table 2. Coverage probability of the HPD credible interval for the CV across sample sizes for 8 combinations of specified values of CV

Sample Size	Highest Posterior Density (HPD)			
	# of times Coverage probability is maintained		Average length	
	<i>Right invariant prior</i>	<i>Jeffrey's prior</i>	<i>Right invariant prior</i>	<i>Jeffrey's prior</i>
10	0	0	*	*
20	0	0	*	*
40	7	0	0.8344	*
60	7	2	0.7933	0.8164
80	6	6	0.3071	0.3009
100	8	8	0.5684	0.5563
150	8	7	0.4562	0.5109
200	8	7	0.3899	0.4382
Overall	44	30	0.5582	0.5244

* Whenever coverage probability is not maintained average length has not been calculated

The posterior density of η is gamma and thus the plot of the density function is smooth. From the histogram and frequency curve it becomes clear that the frequency curve needs to be smoothened at the tail areas. This type of smoothing does not affect the length of the HPD credible interval, but increases the length of the equi-tailed credible interval. This is the reason why the equi-tailed credible intervals are marginally shorter than the HPD credible interval. To incorporate any type of smoothing of a frequency curve in a simulation study is computationally prohibitive and is not attempted here. Figures 1 to 4 represent the posterior density of η and the histogram obtained from 10,000 simulated values of the distribution of $(e^{\sigma^2} - 1)^{\frac{1}{2}}$, corresponding to $n = 60$ and 100, for left and right invariant priors and the value of S_z^2 is fixed at 0.0862 for CV=0.3.

An attempt is also made to study the effect of specified value of CV on the length of credible/confidence interval. Table 3 presents the average length of the interval for various values of CV. From the table it becomes clear that the average length increases as the CV increases for the credible/confidence intervals. The length of the credible interval for the sample size $n=100$, a large value of CV=2.5, for HPD credible interval using right invariant prior is 1.7358 and using Jeffrey's prior is 1.6924 and for confidence interval it is 1.8445. For equi-tailed credible interval for right invariant and Jeffrey's prior it is 1.6747 and 1.6338. The difference in the average length of the confidence interval when CV=0.1 and 2.5, is minimum for equi-tailed credible interval using Jeffrey's prior and is maximum for confidence interval based on M.L.E. The difference in average length for the HPD credible

BAYESIAN INFERENCE FOR VOLATILITY OF STOCK PRICES

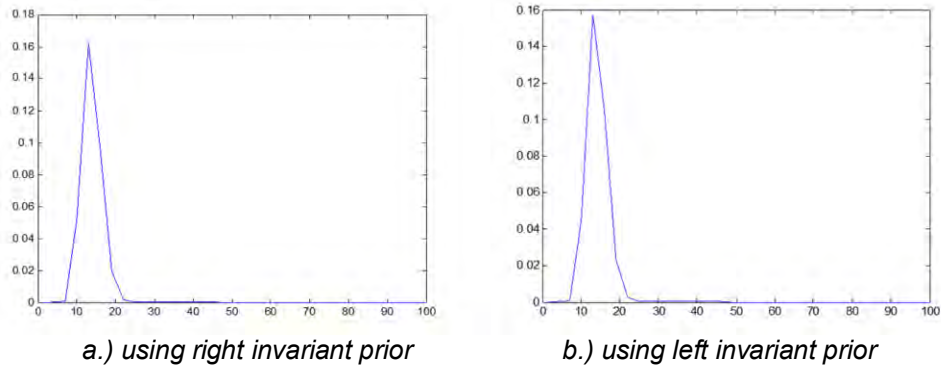


Figure 1. Posterior density of η when $n = 60$

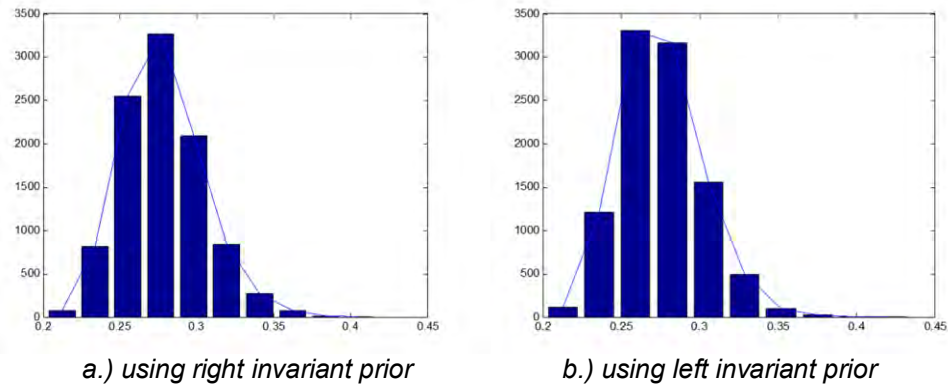


Figure 2. Histogram for $(e^{(1/\eta)} - 1)^{1/2}$ for $n = 60$

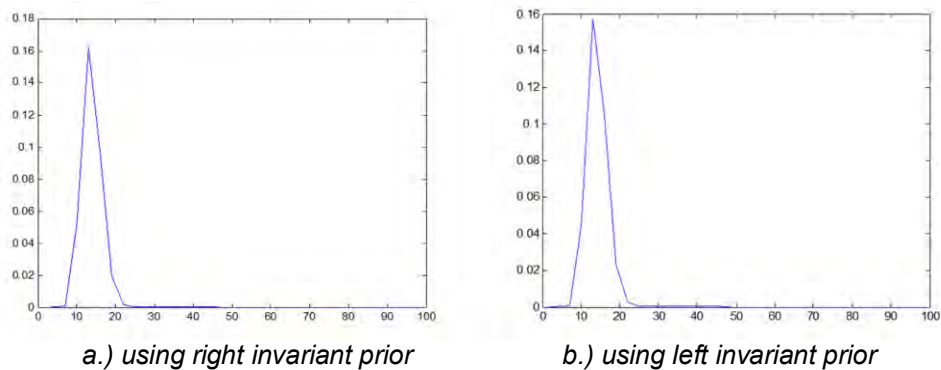


Figure 3. Posterior density of η when $n = 100$

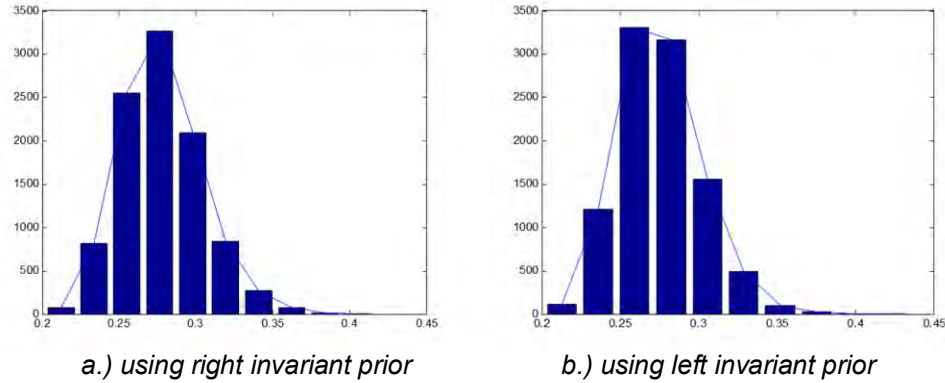


Figure 4. Histogram for $(e^{(1/n)} - 1)^{1/2}$ for $n = 100$

interval based on right invariant and left invariant priors are 1.7080 and 1.6649. The same pattern can be observed for other sample sizes. The average length of HPD credible interval for Jeffrey's prior is marginally higher compared to right invariant prior for all sample sizes and all values of CV under consideration. The coverage probability for these two priors indicates that the coverage probabilities are nearly the same. From the objective Bayesian analysis it amounts to the fact that Bayes procedure is robust against the specification of right and left invariant priors.

Table 3. Average length of the credible and confidence intervals for various values of CV when the sample size is $n = 100$.

Type of interval	Average length when CV equal to								Range
	0.1	0.3	0.5	0.7	1	1.5	2	2.5	
Equi-tailed credible interval with right invariant prior	0.0274	0.0858	0.1527	0.2322	0.3848	0.7153	1.1378	1.6747	1.6473
Equi-tailed credible interval with left invariant prior	0.0271	0.0848	0.1508	0.2337	0.3876	0.7205	1.1138	1.6338	1.6067
Confidence interval based on M.L.E	0.0284	0.0891	0.1593	0.2438	0.4077	0.7689	1.2393	1.8445	1.8161
HPD credible interval with right invariant prior	0.0278	0.0872	0.1554	0.2368	0.3937	0.7354	1.1748	1.7358	1.7080
HPD credible interval with left invariant prior	0.0275	0.0862	0.1534	0.2337	0.3876	0.7205	1.1489	1.6924	1.6649

Analysis of Stock Prices

The advantage of Bayesian analysis is that one can constantly upgrade their knowledge regarding the parameter. This is helpful for making future prediction. In this example the Bayes estimation of the index volatility per mean return is discussed with respect to the stock prices of 3 scripts belonging to large cap category, namely RELIANCE, ACC and TATASTEEL, of the Indian stock market. The daily data from August 19 to November 6, 2013 is used in this analysis. Starting with one week daily data as the training set, Bayes credible interval is obtained for the volatility per mean return. Subsequently the Bayes estimator for successive weeks is computed and the process is continued till the week for which the Bayes estimator lies outside the credible interval. The exercise is repeated with various starting weeks. Table 4 summarizes these results.

Table 4. Bayes credible interval for the index volatility per mean return based on 1 week data and the Bayes estimator for the successive weeks for different starting values.

Stock	Starting Value	95% credible interval	Bayes Estimator		
			2 nd week	3 rd week	4 th week
RELIANCE	Sept 17 th -Sept 23 rd	[0.0877,0.2714]	0.1507	0.1486	0.1460
ACC		[0.0396,0.1226]	0.0987	0.1219	0.1425
TATASTEEL		[0.0124,0.0384]	0.0265	0.0821	0.0820
RELIANCE	Oct 1 st - Oct 8 th	[0.0865,0.2745]	0.1460	0.1421	0.1103
ACC		[0.0713,0.2265]	0.1425	0.1570	0.1704
TATASTEEL		[0.0482,0.1486]	0.0820	0.1381	0.0164

Table 4 shows that based on one week data, the index for the subsequent week for all the three stocks can be accurately predicted. This is true regardless of the starting date namely August 19, September 17, October 1, etc. The duration of the data for making future predictions was also examined. For this purpose credible intervals were constructed using the first 2 through 10 weeks of data. To save space the results are not reported here. From these results it follows that by increasing the length of the data one do not get much accurate prediction for the successive week. Therefore it may be concluded that minimum data of one week be used for making prediction regarding volatility of the stock prices. If the duration increases, then the volatility increases thereby decreasing the decision of the future forecast.

Subjective Bayesian Analysis

As pointed out previously the advantage of Bayesian analysis is that the decision maker can use his belief for making future prediction. In the present scenario this can be achieved using conjugate prior. In the case of lognormal distribution, the conjugate prior is gamma for the scale parameter $\eta=1/\sigma^2$ where μ is fixed. Thus using Uniform prior for μ , the posterior distribution turns out to be gamma and one can use the program developed in this paper for carrying out subjective Bayesian analysis. The mean and variance of the posterior gamma density is given by $\alpha\beta$ and $\alpha\beta^2$ where $\alpha = (n+2)/2$ and $\beta = \frac{1}{2}(n-1)S_z^2$ under right invariant prior. The parameters α and β can be determined by using past information as well as the subjective belief of the decision maker. The posterior density of the previous week can be used as the prior density for the week under consideration. In addition, the investigator can use his belief to modify the parameters of the posterior density of the previous week. Using past data, this type of subjective Bayesian analysis cannot be carried out and is not attempted in this paper.

Conclusion

This paper concentrates on the Bayesian estimation of the index, namely volatility per mean return. This is a frequently used indicator in the analysis of stock market data. The investigation indicates that Bayes credible intervals have smaller width compared to the confidence interval based on maximum likelihood estimator. Frequentist comparison of the credible interval and confidence interval in terms of coverage probability is not well accepted among the Bayesians. The results of this study support the view that accurate prediction can be made based on a small sample size of $n = 5$ for the volatility per mean return of stock prices. Caution has to be exercised for interpreting the width of the credible/confidence interval. For example if the width increases or decreases by 0.05, this amounts to a percentage change of 25% when $CV = 0.2$. Therefore one should not conclude that the difference in the average length of the credible interval and confidence interval is only marginal. The purpose of this paper is to demonstrate the utility of Bayesian inference for forecasting the stock prices.

This paper derives Bayes estimator and the associated credible intervals for the CV of the lognormal distribution. Lognormal distribution has applications in many areas like reliability studies and survival analysis where the focus is the duration of the lifetime. Although emphasis is given to the estimation of mean and median lifetime, the effectiveness of any treatment regime lies in the control of

variability in duration of lifetime. The results developed in this paper can also be used by researchers in these areas. Lognormal distribution is also used in the analysis of rainfall data (Ananthakrishnan & Soman, 1989) and the primary concern is the variability in rainfall, which is commonly measured using coefficient of variation. In these areas the data can be analyzed using objective Bayesian analysis of CV developed in this paper. Numerical analysis is carried out by writing programs using MATLAB software version 7.0 and can be obtained from the first author.

Acknowledgements

The first author would like to thank Government of India, Ministry of Science and Technology, Department of Science and Technology, New Delhi, for sponsoring her with an INSPIRE fellowship, which enables her to carry out the research program which she has undertaken. She is much honored to be the recipient of this award.

References

- Ahmed, S. E. (1995). A pooling methodology for coefficient of variation. *Sankya, Series B*, 57(1), 57-75.
- Ananthakrishnan, R., & Soman, M. K. (1989). Statistical distribution of daily rainfall and its association with the co-efficient of variation of rainfall series. *International Journal of Climatology*, 9, 485-500.
- Bennett, B. M. (1976). On an approximate test for homogeneity of coefficients of variation. In: W. J. Ziegler, (Ed.), *Contribution to Applied Statistics*, (169-171). Stuttgart: Birkhäuser.
- Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis* (2nd Ed.). New York: Springer-Verlag.
- Breunig, R. (2001). An almost unbiased estimator of the co-efficient of variation. *Economics Letters*, 70(1), 15-19.
- Brief, R. P., & Owen, J. (1969). A note on earnings risk and the co-efficient of Variation. *Journal of Finance*, 24, 901-904.
- De, P., Ghosh, J. B., & Wells, C. E. (1996). Scheduling to minimize the co-efficient of variation. *International Journal of Production Economics*, 44, 249-253.

Doornbos, R., & Dijkstra, J. B. (1983). A multi sample test for the equality of co-efficient of variation in normal population. *Communication in Statistics – Simulation and Computation*, 12, 147-158.

Ghosh, J. K., Delampady, M., & Samanta, T. (2006). *An introduction to Bayesian analysis: Theory and Methods*. New York: Springer.

Gupta, C. R., & Ma, S. (1996). Testing the equality of co-efficient of variation in 'k' normal populations. *Communication in Statistics – Theory and Methods*, 25, 115-132.

Jobson, J. D., & Korkie, B. M. (1981). Performance hypothesis testing with the sharpe and treynor measures. *Journal of Finance*, 36, 889-908.

Liu, W., Pang, W. K., & Huang, W. K. (2006). Exact confidence bounds for the co-efficient of variation of a normally distributed population. *International Journal of Statistics and Systems*, 1(1), 81-86.

McKay, A. T. (1932). Distribution of the co-efficient of variation and the extended *t*-distribution. *Journal of the Royal Statistical Society*, 95, 696-698.

Memmel, C. (2003). Performance hypothesis testing with the Sharpe ratio. *Finance Letters*, 1, 21-23.

Mohmoudvand, R., & Hassani, H. (2009). Two new confidence intervals for the CV in a normal distribution. *Journal of Applied Statistics*, 36(4), 429-442.

Nairy, K. S., & Rao, K. A. (2003). Tests of coefficient of variation in normal population. *Communication in Statistics – Simulation and Computation*, 32(3), 641-661.

Panichkitkosolkul, W. (2009). Improved confidence intervals for a co-efficient of variation of a normal distribution. *Thailand Statistician*, 7(2), 193-199.

Shafer, N. J. & Sullivan, J. A. (1986). A simulation study of a test for the equality of co-efficient of variation. *Communication in Statistics – Simulation and Computation*, 15, 681-695.

Verrill, S. P., Johnson, R. A., & Forest Products Laboratory (U.S.). (2007). *Confidence bounds and hypothesis tests for normal distribution coefficients of variation*. Madison, WI: U.S. Dept. of Agriculture, Forest Service, Forest Products Laboratory.

Local Bandwidths for Improving Performance Statistics of Model-Robust Regression 2

Efosa Edionwe
University of Benin
Benin City, Nigeria

Julian I. Mbegbu
University of Benin
Benin City, Nigeria

Model-Robust Regression 2 (MRR2) method is a semi-parametric regression approach that combines parametric and nonparametric fits. The bandwidth controls the smoothness of the nonparametric portion. We present a methodology for deriving data-driven local bandwidth that enhances the performance of MRR2 method for fitting curves to data generated from designed experiments.

Keywords: Semi-parametric methods, Model-Robust Regression, response surface methodology, local bandwidths

Introduction

The understanding of any system or process is enhanced by the availability of fairly accurate mathematical relations connecting the explanatory variables and the dependent variables (responses) of the system. The desire to obtain such mathematical relations led to the development of response surface methodology (RSM) which is a collection of mathematical and statistical techniques employed for modeling and analysis of problems in which a response of interest is influenced by several explanatory variables (Montgomery, 1999; Wu & Hamada, 2000; Raissi & Farsani, 2009). The objective of RSM is to optimize one or more responses, which are influenced by several explanatory variables.

RSM consists of three main phases, namely the experimental design phase, the modeling phase, and the optimization phase (Del Castillo, 2007). However, the efficiency and reliability of the optimal solutions achieved at the optimization phase depends on the results obtained in the modeling phase. Better results obtained in

Mr. Edionwe is a Ph.D student in the Department of Mathematics. Email him trustnelson24@yahoo.com. Dr. Mbegbu is an Associate Professor in the Department of Mathematics. Email him at julian.mbegbu@yahoo.com.

the modeling phase ensure better optimal solution in the optimization phase (Pickle, 2006).

The modeling phase involves the use of regression techniques to fit a curve to the data generated from the experiment. Regression techniques employed in RSM include parametric regression, nonparametric regression and semi-parametric regression (Pickle, 2006; Wan, 2007).

In Parametric regression, a low-order polynomial such as the k -factor second-order model of the form

$$y_i = \beta_0 + \sum_{i=1}^k \beta_i x_i + \sum_{i=1}^k \beta_{ii} x_i^2 + \sum_{i>j} \beta_{ij} x_i x_j + \varepsilon_i \quad (1)$$

is assumed for fitting the data, where β_i , β_{ii} and β_{ij} are the model parameters, x_i and x_j are the explanatory (Del Castillo, 2007). For n -sample observations, (1) can be expressed in matrix form as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2)$$

where \mathbf{Y} is an n by 1 vector of responses, \mathbf{X} is an n by $\left(1 + 2k + \binom{k}{2}\right)$ model matrix,

$\boldsymbol{\beta}$ is a $\left(1 + 2k + \binom{k}{2}\right)$ by 1 vector of unknown model parameters, and $\boldsymbol{\varepsilon}$ is an n by 1

vector of random errors. The Ordinary Least Squares (OLS) method gives the vector of estimated responses is given as

$$\hat{\mathbf{y}}^{(OLS)} = \mathbf{X}\hat{\boldsymbol{\beta}}_{OLS} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = \mathbf{H}^{(OLS)}\mathbf{y} \quad (3)$$

A disadvantage of the parametric regression method is that if the assumed model is misspecified, the fitted curve is affected by high bias (Einsporn & Birch, 1993; Mays, 2001b; Pickle, 2006).

In nonparametric regression, the user approaches the problem without assuming a model and attempts to fit a curve to the data points by employing a weighting scheme (Uysal & Guvenir, 1999; DiNardo & Tobias, 2001). Most often, nonparametric regression (for example, the kernel regression, local linear regression) is employed when a theoretical reference curve is unavailable for a process and the data size is large (Hens, 2005; Hernández-Lobato, 2010). The local

IMPROVING PERFORMANCE STATISTICS OF MRR2

linear regression (LLR) utilizes kernel weights for smoothing. For instance, the LLR estimate of observation y_0 obtained at the location $\tilde{\mathbf{x}}_0$ is given as

$$\hat{y}_0^{(LLR)} = \tilde{\mathbf{x}}_0' (\tilde{\mathbf{X}}' \mathbf{W}_0 \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}' \mathbf{W}_0 \mathbf{y} = \mathbf{h}_0^{(LLR)} \mathbf{y} \quad (4)$$

$$\tilde{\mathbf{X}} = \begin{bmatrix} \tilde{\mathbf{x}}_1' \\ \tilde{\mathbf{x}}_2' \\ \vdots \\ \tilde{\mathbf{x}}_n' \end{bmatrix} \quad (5)$$

where $\tilde{\mathbf{x}}_i' = (1 \ x_{i1}, \dots, x_{ik})$, the n by n diagonal matrix \mathbf{W}_0 , known as the local weight matrix for location \mathbf{x}_0 , is given by

$$\mathbf{W}_0 = \text{diag} \left(h_{01}^{(KER)} \quad h_{02}^{(KER)} \quad \dots \quad h_{0n}^{(KER)} \right) \quad (6)$$

$h_{0i}^{(KER)}$ represents a kernel weight assigned to y_i in the estimation of y_0 at location \mathbf{x}_0 and is given as

$$h_{0i}^{(KER)} = \frac{K\left(\frac{x_0 - x_i}{b}\right)}{\sum_{i=1}^n K\left(\frac{x_0 - x_i}{b}\right)} \quad (7)$$

K is a univariate kernel function, b is referred to as the bandwidth (Härdle, 1994). A commonly used kernel function is the simplified Gaussian kernel function given as

$$K\left(\frac{x_0 - x_i}{b}\right) = e^{-\left(\frac{x_0 - x_i}{b}\right)^2} \quad (8)$$

For the multivariate case with k explanatory variables, a common form of the Gaussian kernel function used is the product kernel given as

$$K(\tilde{x}_0, \tilde{x}_1) = \prod_{j=1}^k K\left(\frac{\tilde{x}_{0j} - \tilde{x}_{1j}}{b}\right) \quad (9)$$

$\tilde{\mathbf{x}}_0 = (x_{01}, x_{02}, \dots, x_{0k})$ is the prediction point, K is the univariate kernel function (Wan, 2007). In general, using matrix notation, LLR estimated response can be written as

$$\hat{\mathbf{y}}^{(LLR)} = \mathbf{H}^{(LLR)} \mathbf{y} \quad (10)$$

where $H^{(LLR)}$ is the LLR ‘‘HAT’’ or smoother matrix defined as

$$H^{(LLR)} = \begin{bmatrix} \mathbf{h}_1^{(LLR)'} \\ \mathbf{h}_2^{(LLR)'} \\ \vdots \\ \mathbf{h}_n^{(LLR)'} \end{bmatrix} \quad (11)$$

$$\mathbf{h}_i^{(LLR)'} = \tilde{\mathbf{x}}_i' (\tilde{\mathbf{X}}' \mathbf{W}_i \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}' \mathbf{W}_i \quad (12)$$

A disadvantage of nonparametric methods is that large amounts of data are required. Moreover, the capacity of nonparametric methods to describe complex patterns makes them more prone to overfitting (Mays, 2001b; Wan, 2007; Starnes, Birch & Robinson, 2008).

Semiparametric regression methods involve fitting the data both parametrically and nonparametrically, and then combining the results to form a curve that is based on suitable theoretical form, yet still being able to adapt to aberrations or misspecifications from that form. Hence semi-parametric regression techniques are robust to model misspecifications (Starnes, 1999; Mays, 2001a; Hens, 2005).

Starnes (1999) and Pickle (2006) reported that MRR2 is the best overall semi-parametric regression procedure for fitting small-sample data in situation of small to moderate model misspecification.

IMPROVING PERFORMANCE STATISTICS OF MRR2

MRR2 technique combines a parametric fit like $\hat{\mathbf{y}}^{(OLS)}$ to the raw data and a nonparametric fit to the vector of residuals, \mathbf{r} , from the parametric fit. The MRR2 fit at location \mathbf{x}_0 is given by

$$\hat{y}_0^{(MRR2)} = \tilde{\mathbf{x}}_0' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} + \lambda \tilde{\mathbf{x}}_0' (\tilde{\mathbf{X}}'\mathbf{W}_r\tilde{\mathbf{X}}')^{-1} \tilde{\mathbf{X}}'\mathbf{W}_r (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}') \mathbf{y} \quad (13)$$

where

$$\mathbf{h}_0^{(MRR2)'} = \mathbf{x}_0' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' + \lambda \tilde{\mathbf{x}}_0' (\tilde{\mathbf{X}}'\mathbf{W}_r\tilde{\mathbf{X}}')^{-1} \tilde{\mathbf{X}}'\mathbf{W}_r (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'),$$

\mathbf{r} is the vector of residuals from the parametric fits, \mathbf{I} is an n by n identity matrix, \mathbf{W}_r is the n by n diagonal matrix containing the kernel weights for fitting the parametric residuals and is obtained using the same procedure as in (6), the matrices \mathbf{X} and $\tilde{\mathbf{X}}$ are as defined in (2) and (5) respectively. MRR2 fits is expressed in matrix form as

$$\hat{\mathbf{y}}^{(MRR2)} = \left[\mathbf{H}^{(OLS)} + \lambda \mathbf{H}^{(LLR)} (\mathbf{I} - \mathbf{H}^{(OLS)}) \right] \mathbf{y} \quad (14)$$

$$= \mathbf{H}^{(MRR2)} \mathbf{y} \quad (15)$$

where $\mathbf{H}^{(MRR2)}$ is the MRR2 “HAT” matrix defined as

$$\mathbf{H}^{(MRR2)} = \begin{bmatrix} \mathbf{h}_1^{(MRR2)'} \\ \mathbf{h}_2^{(MRR2)'} \\ \vdots \\ \mathbf{h}_n^{(MRR2)'} \end{bmatrix} \quad (16)$$

Wan (2007) reported two expressions for the mixing parameter, λ . One is the estimated asymptotically optimal mixing parameter given by

$$\hat{\lambda}^* = \frac{\langle \hat{r}, r \rangle}{\|\hat{r}\|^2} \quad (17)$$

the notation $\langle \rangle$ represents the inner product and $\| \cdot \|$ is the standard L_2 (Euclidean) norm. A second data-driven method chooses $\hat{\lambda}^*$ such that $PRESS^{**}(\lambda)$ defined for a given optimal bandwidth, b^* , as

$$PRESS^{**}(\lambda) = \frac{\sum_{i=1}^n \left(y_i - \hat{y}_{i,-1}(b^*, \lambda) \right)^2}{n - \text{tr} \left(H^{(MRR2)}(b^*, \lambda) \right) + (n - k - 1) \frac{SSE_{\max} - SSE_{b^*}}{SSE_{\max}}} \quad (18)$$

is minimized. n is the sample size, b^* is the optimal bandwidth, k is the number explanatory variables, SSE_{\max} is the maximum sum of squared errors obtained as b tends to infinity, SSE_{b^*} is the sum of squared errors for the optimal bandwidth, $\text{tr} \left(H^{(MRR2)}(b^*, \lambda) \right)$ is the trace of the MRR2 “HAT” matrix for a given b^* and λ , and $\hat{y}_{i,-1}$ is the fit at x_i with the i^{th} observation left out.

The bandwidth is an important parameter in that it determines performance of the model in terms of criteria such as variance, mean squared error (Huang & Fan, 1996). A bandwidth is said to be fixed or global if its value is constant for the full range of the data or if does not change with locations or runs in a given regression procedure otherwise it is referred to as local, variable or adaptive bandwidth. For a given location, local bandwidths are chosen according to factors involving the values of the explanatory variables, x_i , or of the response, y_i , or both (Starnes, 1999). This dependence allows different degree of smoothing for different locations in the data thereby giving the data more privilege to determine the functional form of the model fitted and to incorporate the information provided by the density of the data.

Among the categories of methods for selecting bandwidths, the most frequently employed procedures include the plug-in methods and the cross-validation methods (Fan & Gijbels, 1992; Atkeson, Moore & Schaal, 1997; Gerard & Schucany, 1999; Racine, 2008; Avery, 2010; Kohler, Schindler and Sperlich, 2011). However, all the criteria for selecting bandwidths are based on the same philosophy, and they are such that the fitted value $\hat{y}(x)$ is as close to the true value $y(x)$ as possible thereby minimizing errors associated with estimation (Härdle,

1994; Galdo & Black, 2008). Researches applying MRR2 for fitting curves to RSM data employ a penalized version of prediction error sum of squares referred to as $PRESS^{**}$ and given by

$$PRESS^{**}(b) = \frac{\sum_{i=1}^n (y_i - \hat{y}_{i,-1}(b))^2}{n - tr(H^{(MRR2)}(b)) + (n-k-1) \frac{SSE_{\max} - SSE_b}{SSE_{\max}}} \quad (19)$$

$$PRESS^{**}(b) = \sum_{i=1}^n (y_i - \hat{y}_{i,-1}(b))^2, i = 1, 2, \dots, n \quad (20)$$

where all the parameters retain their previous definitions in (18) (Pickle, 2005; Wan, 2007).

The remainder of this paper is organized as follows: A review of methods for deriving local bandwidth is presented in the next section. A new methodology for deriving data-driven local bandwidths follows. After that, results of application of MRR2 and LLR methods utilizing the new data-driven local bandwidths to a multiresponse problem are presented. Finally, a discussion on the comparison of results from OLS, LLR and MRR2 (both fixed optimal bandwidth and local optimal bandwidths) is presented.

A Review of Methods for Deriving Local Bandwidths

For kernel density estimation, Fan and Gijbels (1992), using Average Mean Integrated Squared Error, gave an expression for optimal variable bandwidths as

$$\alpha_{opt}(x) = \begin{cases} b \left(\frac{f_X(x) [m''(x)]^2}{\sigma^2(x)} \right)^{\frac{1}{5}}, & \text{if } W(x) > 0, \\ \alpha^*(x), & \text{if } W(x) = 0 \end{cases} \quad (21)$$

where b is any arbitrarily positive constant, $\alpha^*(x)$ can take any value greater than zero, $W(x)$ is a diagonal matrix of weights, $m''(x)$ is the second-derivative of the unknown function, $f_X(\cdot)$ and $\sigma^2(\cdot)$ are the marginal density of X and the conditional variance of Y given X respectively. The limitation in the use of this variable bandwidth is that it requires estimates of $f_X(\cdot)$, $m''(x)$, and $\sigma^2(\cdot)$ respectively. Hence, the efficiency of (21) depends on how close these estimates are to the true values.

Schucany (1995) proposed a variable bandwidth selector for both the kernel and local linear regression. An expression for the optimal bandwidth is given by

$$h_{opt}^{SCH}(x) = \left(\frac{\sigma^2 A}{2pnB(x)^2} \right)^{1/(2p+1)} \quad (22)$$

where p is the degree of the polynomial, n is the number of observations, A is a constant which depends on the kernel, $B(x)$ is an approximation for the bias. Again, $h_{opt}^{SCH}(x)$ is calculated using estimates of σ^2 and $B(x)$. Hence, the quality of the final estimator \hat{h}_{opt}^{SCH} depends on the choice of a “pilot bandwidth” from which an estimate of the $B(x)$ is obtained. Moreover, (22) is developed for cases where the levels of a single explanatory variable are equally-spaced.

Few of the plug-in methods for obtaining variable bandwidths are used in practice due to computational difficulty. Plugs methods seem logically inconsistent since they require higher order smoothness of the unknown function (Bickels & Li, 2007; Galdo & Black., 2008; Avery, 2010).

A Local cross-validation variable bandwidth which reflects the impact of the responses and suitable for a single explanatory is considered in Zheng (2010) and is given as

$$h^{**}(x) = \arg \left\{ \min_h \sum_{i=1}^{l(x)} \left(Y'_i - \hat{m}_{-i}(X'_i) \right)^2 \right\} \quad (23)$$

where $l(x)$ denotes the number of covariate values falling in a certain defined interval $I_{x=[x-h^*(x), x+h^*(x)]}$ and $(X'_i, Y'_i), i=1, \dots, l(x)$ denotes the number observations falling in the interval, $h^*(x)$ is a sequence of a version of optimal Bayesian bandwidths, and $\hat{m}_{-i}(X_i)$ is given as

$$\hat{m}_{-i}(X_i) = \frac{1}{(n-1)h} \sum_{j \neq i} K \left(\frac{x_i - x_j}{h} \right) \frac{M_{2i}(x_i) - \left(\frac{x_i - x_j}{h} \right) M_{1i}(x_i)}{M_{2i}(x_i) M_{0i}(x_i) - M_{1i}^2(x_i)} Y_j \quad (24)$$

and

$$M_{ci}(X_i) = ((n-1)h)^{-1} \sum_{j \neq i} K\left(\frac{x_i - x_j}{h}\right) \left(\frac{x_i - x_j}{h}\right)^c, \text{ for } c = 1, 2, 3 \quad (25)$$

The method in (24) works for a single explanatory variable. Besides, the choice of $l(x)$ is dependent on $h^*(x)$ which, according to the author, requires estimates of some prior parameters.

A methodology for the derivation of a function for generating local bandwidth is presented in the section that follows. The local bandwidth generated by the function can be applied to data with more than one explanatory variable. Furthermore, typical of cross validation procedures, no estimates of parameters is required for the utilization of the proposed function.

Methodology

Derivation of a Function for Generating Local Bandwidth

A new methodology used to derive a function for generating data-driven local bandwidth is presented. In deriving the function, the basic objectives to achieve are: to allow the values of the bandwidths to be a function of the observations we intend to fit; to assume that a real number N , which also acts as a tuning parameter is the sum of all the bandwidths that minimize $PRESS^{**}$. The simplified kernel function, which is a decreasing function, is utilized in the paper. Therefore, the function generating the local bandwidth is modeled in a manner that locations with relatively smaller observations are assigned smaller bandwidths (corresponding to heavier weights via the kernel function), and vice versa. For convenience, this function is referred to as N -squared function and its derivation is as follows:

Given that $T = \sum_{i=1}^n y_i$, (T is the sum of all the observations), n is the number of observations or locations, or sample size, b_i , $i = 1, \dots, n$, is the bandwidth for the i^{th} location and N is the sum of the bandwidths that minimize $PRESS^{**}$.

First, it is required that for each location, the bandwidth be a function of the size of observation at that location, and one of the ways to achieve this is to express the bandwidth, b_i , as ratio of the i^{th} observation to the sum of the data, T

$$b_i = \frac{y_i}{T}, \text{ for } i = 1, \dots, n. \quad (26)$$

Because the simplified Kernel function is a decreasing function, hence to ensure heavier kernel weights are assigned to smaller observations and vice versa, smaller observations need to have smaller bandwidths and to achieve this, (26) is expressed as

$$b_i = \left(N - \frac{y_i}{T} \right) \quad (27)$$

Taking sum of both sides of (27) gives

$$\sum_i^n b_i = \sum_1^n \left(N - \frac{y_i}{T} \right) \quad (28)$$

$$\sum_i^n b_i = \sum_1^n \left(\frac{NT - y_i}{T} \right) \quad (29)$$

Next, proceed to determine a factor that multiplies the right-hand side of (29) to ensure the bandwidths sum to a value N .

By expanding the right side of (29)

$$\sum_i^n b_i = \left(\frac{(NT - y_1) + (NT - y_2) + \dots + (NT - y_n)}{T} \right) \quad (30)$$

On collecting like terms in (30)

$$\sum_i^n b_i = \frac{NTn - \sum_{i=1}^n y_i}{T} \quad (31)$$

But the sum of the data, $\sum_{i=1}^n y_i$, is equal to T as previously defined hence (31) reduces to

$$\sum_i^n b_i = \frac{NTn - T}{T} \quad (32)$$

Therefore

$$\sum_i^n b_i = (Nn - 1) \quad (33)$$

which implies $\sum_i^n \left(\frac{NT - y_i}{T} \right) = (Nn - 1)$. Hence, to ensure the bandwidths sum to a value N , we need to multiply the right hand side of (29) by a factor $N / (Nn - 1)$, giving

$$\sum_i^n b_i = \sum_i^n \left(\frac{NT - y_i}{T} \right) N / (Nn - 1) \quad (34)$$

$$\sum_i^n b_i = \sum_i^n \frac{N(NT - y_i)}{T(Nn - 1)} \quad (35)$$

Therefore

$$b_i = \frac{N^2 T - N y_i}{T(Nn - 1)}, \quad i = 1, 2, \dots, n, \quad (36)$$

$$b_i = \frac{N^2 \sum_{i=1}^n y_i - N y_i}{(Nn - 1) \sum_{i=1}^n y_i}, \quad i = 1, 2, \dots, n. \quad (37)$$

Equation (37) gives the N-squared function for data-driven variable bandwidths. The optimal local bandwidth, \mathbf{b} is a vector whose elements are the bandwidths b_i , (for smoothing i^{th} location of the observation), $i = 1, 2, \dots, n$, obtained at the value of N in (37) where $PRESS^{**}(\mathbf{b})$ given by

$$PRESS^{**}(\mathbf{b}) = \frac{\sum_i^n (y_i - y_{i,-i}(\mathbf{b}))^2}{n - \text{trace}(H^{(\cdot)}(\mathbf{b})) + (n - k - 1) \frac{(SSE_{\max} - SSE_{\mathbf{b}})}{SSE_{\max}}} \quad (38)$$

$$= \frac{PRESS(\mathbf{b})}{n - trace(H^{(\cdot)}(\mathbf{b})) + (n - k - 1) \frac{(SSE_{\max} - SSE_{\mathbf{b}})}{SSE_{\max}}} \quad (39)$$

where $H^{(\cdot)}$ is the “HAT” matrix for (LLR) or MRR2 obtained by using local bandwidths from N-squared function, SSE_{\max} is the maximum sum of squares of errors over all possible bandwidths which is equivalent to $\sum_{i=1}^n (y_i - \hat{y}_i^{(OLS)})^2$ for LLR or $\sum_{i=1}^n (e_i - \hat{e}_i^{(OLS)})^2$ for MRR2 where $\hat{\mathbf{y}}^{(OLS)}$ and $\hat{\mathbf{e}}^{(OLS)}$ are the OLS fit of a first-order model for responses and OLS residuals respectively, SSE_{b_i} is given by $\sum_{i=1}^n (y_i - \hat{y}_i^{(LLR)}(b_i))^2$ for LLR or $\sum_{i=1}^n (y_i - \hat{y}_i^{(MRR2)}(b_i))^2$ for the MRR2 counterpart. (See Wan, 2007). For MRR2, the mixing parameter, λ , is obtained using equations (17) or (18).

LLR and MRR2 methods are applied using local bandwidth derived from N-squared function to the Minced Fish Quality problem Wan (2007) and its performance is compared with results from parametric, (OLS), LLR, (fixed bandwidth), and MRR2, (fixed bandwidth), approaches. The comparison is based on some performance criteria including, estimate of the variance, (S^2), the coefficient of determination, (R^2), adjusted coefficient of determination, (R_{adj}^2), PRESS given in (20), $PRESS^* = PRESS/DF_{error}$, where $DF_{error} = DF_{total} - DF_{model}$, and $PRESS^{**}$.

Application of Local Bandwidths from N-Squared Function

The data for the Minced Fish Quality problem presented in Wan (2007) is from food science and is used here to compare the performance of the modeling techniques discussed herein. The problem involves three independent variables x_1 , x_2 , x_3 which represent washing temperatures, washing time, washing ratio of water volume to sample weight respectively, and four response variables y_1 , y_2 , y_3 , y_4 , representing springiness, thiobarbituric acid number, (TBA), cooking loss, and whiteness index respectively. The observed data were collected through a Central Composite Design, (CCD), and is presented in Table 1.

According to Wan (2007), the final fitted second-order models for OLS for responses y_1 and y_4 include three terms: intercept, x_1 and x_1^2 . The OLS model for response y_2 includes five terms: intercept, x_1 , x_2 , x_1^2 , and x_{12} . The OLS model for

IMPROVING PERFORMANCE STATISTICS OF MRR2

response y_3 has eight terms: intercept, x_1 , x_2 , x_3 , x_1^2 , x_{12} , x_{13} , and x_3^2 . Therefore, the model spaces for the OLS part of the MRR2 are the same for the final fitted OLS models for each of the responses while the LLR and the nonparametric part of MRR2 utilize a first-order version of the OLS models since LLR and the nonparametric part of MRR2 are based local linear smoothing. Thus, the model matrix for the LLR and the nonparametric part of MRR2 for response y_3 consists of four terms: intercept, x_1 , x_2 , and x_3 , the model matrices for response y_1 and y_4 both consist of the intercept and x_1 , and that for response y_2 consists of the intercept, x_1 and x_2 .

Table 1. A CCD with three factors and four responses on minced fish quality

	Coded Variables			Response values			
	x_1	x_2	x_3	y_1	y_2	y_3	y_4
1	0.203	0.203	0.203	1.83	29.31	29.50	50.36
2	0.797	0.203	0.203	1.73	39.32	19.40	48.16
3	0.203	0.797	0.203	1.85	25.16	25.70	50.72
4	0.797	0.797	0.203	1.67	40.18	27.10	49.69
5	0.203	0.203	0.797	1.86	29.82	21.40	50.09
6	0.797	0.203	0.797	1.77	32.20	24.00	50.61
7	0.203	0.797	0.797	1.88	22.01	19.60	50.36
8	0.797	0.797	0.797	1.66	40.02	25.10	50.42
9	0	0.5	0.5	1.81	33.00	24.20	29.31
10	1	0.5	0.5	1.37	51.59	30.60	50.67
11	0.5	0	0.5	1.85	20.35	20.90	48.75
12	0.5	1	0.5	1.92	20.53	18.90	52.70
13	0.5	0.5	0	1.88	23.85	23.00	50.19
14	0.5	0.5	1	1.90	20.16	21.20	50.86
15	0.5	0.5	0.5	1.89	21.72	18.50	50.84
16	0.5	0.5	0.5	1.88	21.21	18.60	50.93
17	0.5	0.5	0.5	1.87	21.55	16.80	50.98

Table 2. Bandwidths for LLR_{VB1} for each location and response

Location	y_1	y_2	y_3	y_4
1	0.0801	0.4377	0.5355	0.0800
2	0.0803	0.4365	0.5370	0.0801
3	0.0800	0.4382	0.5361	0.0799
4	0.0804	0.4364	0.5358	0.0800
5	0.0800	0.4376	0.5367	0.0800
6	0.0802	0.4373	0.5363	0.0800
7	0.0800	0.4385	0.5370	0.0800
8	0.0804	0.4364	0.5361	0.0800
9	0.0801	0.4372	0.5363	0.0815
10	0.0810	0.4350	0.5353	0.0799
11	0.0800	0.4387	0.5368	0.0801
12	0.0799	0.4387	0.5371	0.0798
13	0.0800	0.4383	0.5365	0.0800
14	0.0799	0.4388	0.5367	0.0799
15	0.0799	0.4386	0.5372	0.0799
16	0.0800	0.4386	0.5371	0.0799
17	0.0800	0.4386	0.5374	0.0799

Table 3. Bandwidths for $MMR2_{VB1}$ for each location and response

Location	y_1	y_2	y_3	y_4
1	0.0792	0.2568	0.3624	0.0791
2	0.0794	0.2556	0.3640	0.0793
3	0.0792	0.2573	0.3630	0.0791
4	0.0796	0.2555	0.3628	0.0792
5	0.0792	0.2567	0.3637	0.0792
6	0.0794	0.2564	0.3633	0.0791
7	0.0791	0.2577	0.3639	0.0791
8	0.0796	0.2555	0.3631	0.0791
9	0.0793	0.2563	0.3632	0.0807
10	0.0802	0.2541	0.3622	0.0791
11	0.0792	0.2579	0.3637	0.0793
12	0.0791	0.2578	0.3640	0.0790
13	0.0791	0.2574	0.3634	0.0792
14	0.0791	0.2579	0.3637	0.0791
15	0.0791	0.2577	0.3641	0.0791
16	0.0791	0.2578	0.3641	0.0791
17	0.0792	0.2577	0.3644	0.0791

Results

The values of SSE_{max} for LLR for y_1, y_2, y_3 , and y_4 are 0.1638, 942.9793, 234.8291, and 352.1950 respectively, and those for $MRR2$ are 0.0231, 90.9033, 41.1338, and 198.8048, respectively. The optimal Local bandwidths for each response generated for a given value of N in N-squared function are given in Table 2, and Table 3 for LLR_{VB1} and $MMR2_{VB1}$ respectively.

Tables 4–6 present the results of numerical values of performance statistics for comparing OLS, LLR for both fixed bandwidth and local bandwidth generated via N-squared function, and $MRR2$ for both fixed bandwidth and local bandwidth generated via N-squared function. For convenience, LLR and $MRR2$ for fixed bandwidth reported in Wan (2007) are referred to as LLR_{FB} and $MMR2_{FB}$ respectively while LLR and $MRR2$ for N-squared variable bandwidths function are designated LLR_{VB1} and $MMR2_{VB1}$ respectively. The case where the values of the mixing parameters for all responses are all 1 is considered for comparison sake. This will enable one attribute the performance of the models solely to the type of bandwidth used rather than to values of the mixing parameters. Best values for each performance statistics and for each response are shown in bold print.

Table 4 compares the performance statistics of fitted responses from the three regression methods discussed here. LLR_{FB} produces best results exclusively in zero cell and joint best result in zero cell, $MMR2_{FB}$ produces best result exclusively in zero cell and joint best results in zero cell. OLS produces best results exclusively in three cells and joint best result in zero cell. LLR_{VB1} produces best results exclusively in six cells and joint best results in six points. $MMR2_{VB1}$ produces best results exclusively in nine cells and joint best results in six cells. $MMR2_{VB1}$ produces the smallest S^2 , highest R^2 and R^2_{adj} exclusively across two of the responses and joint best results for these statistics in the remaining two responses. For DF_{error} , $MMR2_{VB1}$ produces either the best or competitive results across all responses. In addition, $MMR2_{VB1}$ produces competitive results in several cells where it fails to produce the best results. Table 5 compares the performance statistics of fitted responses from the two versions of local linear regression, LLR_{FB} and LLR_{VB1} .

LLR_{FB} produces best results in just five cells in a total of twenty-four cells and LLR_{VB1} produces best results in nineteen cells which is equivalent to 79.17% of the total cells for comparison. Table 6 compares the performance statistics of fitted responses from the two versions of model-robust regression, $MMR2_{FB}$ and $MMR2_{VB1}$. $MMR2_{FB}$ produces best results exclusively in just one cell and joint best results in zero cell. $MMR2_{VB1}$ produces best results exclusively in twenty-three cells and best results in zero cell which is equivalent to 95.83% of the total cells for

comparison. Figures 1 through 4 present the plots of each response against x_1 for various values of x_2 and x_3 were applicable. The $MMR2_{(FB)}$ and $MMR2_{(VB1)}$ overlap in virtually all the plots for y_1 and y_4 reflecting the closeness of several performance statistics for the two approaches. However, in y_2 and y_3 plots, the $MMR2_{(VB1)}$ plots are seen passing through the mean values of the responses for instance plot in the top right in Figure 4.

The results in Tables 4 and 6 clearly show that $MMR2_{VB1}$ is the overall best regression technique outperforming the $MMR2_{FB}$ that produces best results in sixteen cells out of twenty-four cells in results presented in wan (2007) where it is compared with OLS and LLR_{FB} . However, in situations where the user has no prior knowledge of the true underlying model LLR_{VB1} will certainly come in handy as results in Table 4 reveal.

Table 4. Results of comparison of performance statistics of OLS, fixed bandwidth LLR and MRR2, and Variable bandwidth LLR and MRR2 all for $\lambda = 1$, fixed optimal bandwidth, b , and N as defined in equation (37) for local bandwidths in Table 2, and Table 3

	METHOD	b	N	DF_{error}	S^2	R	R^2_{adj}	PRESS	PRESS'	PRESS''
y_1	OLS	-	-	14.0000	1.65E-03	0.9211	0.9090	0.0582	0.0042	0.0042
	LLR_{FB}	0.146		12.1381	1.04E-03	0.9570	0.9433	0.0682	0.0056	0.0026
	$MRR2_{FB}$	0.17		12.2680	1.03E-03	0.9568	0.9436	0.0472	0.0039	0.0025
	LLR_{VB1}		1.362	12.0000	1.00E-03	0.9579	0.9439	0.0216	0.0018	0.0008
	$MRR2_{VB1}$		1.348	12.0000	1.00E-03	0.9579	0.9439	0.0405	0.0034	0.0021
y_2	OLS	-	-	12.0000	7.5417	0.9341	0.9122	234.1166	19.5097	19.5097
	LLR_{FB}	0.436		11.2120	21.8508	0.8217	0.7456	785.7855	70.0873	36.4222
	$MRR2_{FB}$	0.277		8.9400	4.8253	0.9686	0.9438	319.3332	35.7214	19.6311
	LLR_{VB1}		7.441	11.2260	21.9206	0.8209	0.7448	785.9495	70.0115	36.4328
	$MRR2_{VB1}$		4.366	8.6923	4.6819	0.9704	0.9455	305.1765	35.1090	18.5803
y_3	OLS	-	-	9.0000	4.5641	0.8408	0.7170	182.4468	20.2719	20.2719
	LLR_{FB}	0.537		8.3730	9.7990	0.6821	0.3925	287.0564	34.2849	17.0554
	$MRR2_{FB}$	0.542		6.5960	2.9031	0.9258	0.8200	177.6750	26.9357	13.1264
	LLR_{VB1}		9.121	8.3672	9.7791	0.6829	0.3937	286.6772	34.2622	17.0261
	$MRR2_{VB1}$		6.179	3.9265	1.3817	0.9790	0.9143	173.9599	44.3046	11.4358
y_4	OLS	-	-	14.0000	14.2182	0.5407	0.4751	684.7407	48.9101	48.9101
	LLR_{FB}	0.12		12.0310	1.0197	0.9717	0.9624	454.5871	37.7832	17.1484
	$MRR2_{FB}$	0.119		12.0290	1.0158	0.9718	0.9625	486.8458	40.4725	18.6472
	LLR_{VB1}		1.361	12.0000	1.0116	0.9720	0.9627	407.8131	33.9844	15.3990
	$MRR2_{VB1}$		1.347	12.0000	1.0116	0.9720	0.9627	451.5303	37.6275	17.3105

IMPROVING PERFORMANCE STATISTICS OF MRR2

Table 5. LLR_{FB} versus LLR_{VB1} for values of b and b_i indicated in Table 2

	METHOD	DF_{error}	S^2	R	R^2_{adj}	PRESS	PRESS*	PRESS**
y_1	LLR_{FB}	12.1380	1.04E-03	0.9570	0.9433	0.0682	0.0056	0.0026
	LLR_{VB1}	12.0000	1.00E-03	0.9579	0.9439	0.0216	0.0018	0.0008
y_2	LLR_{FB}	11.2120	21.8508	0.8217	0.7456	785.7855	70.0873	36.4222
	LLR_{VB1}	11.2260	21.9206	0.8209	0.7448	785.9495	70.0115	36.4328
y_3	LLR_{FB}	8.3730	9.7990	0.6821	0.3925	287.0564	34.2849	17.0554
	LLR_{VB1}	8.3672	9.7791	0.6829	0.3937	286.6772	34.2622	17.0261
y_4	LLR_{FB}	12.0310	1.0197	0.9717	0.9624	454.5871	37.7832	17.1484
	LLR_{VB1}	12.0000	1.0116	0.9720	0.9627	407.8131	33.9844	15.3990

Table 6. $MMR2_{FB}$ versus $MMR2_{VB1}$ for values of b and b_i indicated in Table 3.

	METHOD	DF_{error}	S^2	R	R^2_{adj}	PRESS	PRESS*	PRESS**
y_1	$MMR2_{FB}$	12.2680	1.03E-03	0.9568	0.9436	0.0472	0.0039	0.0025
	$MMR2_{VB1}$	12.0000	1.00E-03	0.9579	0.9439	0.0405	0.0034	0.0021
y_2	$MMR2_{FB}$	8.9400	4.8253	0.9686	0.9438	319.3332	35.7214	19.6311
	$MMR2_{VB1}$	8.6923	4.6819	0.9704	0.9455	305.1765	35.1090	18.5803
y_3	$MMR2_{FB}$	6.5960	2.9031	0.9258	0.8200	177.6750	26.9357	13.1264
	$MMR2_{VB1}$	3.9265	1.3817	0.9790	0.9143	173.9599	44.3046	11.4358
y_4	$MMR2_{FB}$	12.0290	1.0158	0.9718	0.9625	486.8458	40.4725	18.6472
	$MMR2_{VB1}$	12.0000	1.0116	0.9720	0.9627	451.5303	37.6275	17.3105

Figures 1 and 2 compare the plots of \hat{y}_1 versus x_1 and \hat{y}_4 versus x_1 respectively, using OLS, MRR2 via fixed bandwidth $MMR2_{FB}$, and MRR2 via local bandwidths $MMR2_{VB1}$ from N-squared function.

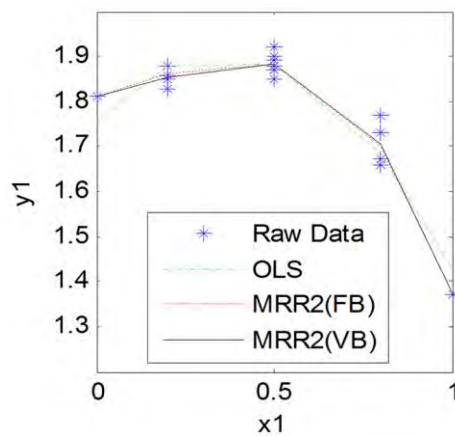


Figure 1. Plot of \hat{y}_1 vs. x_1

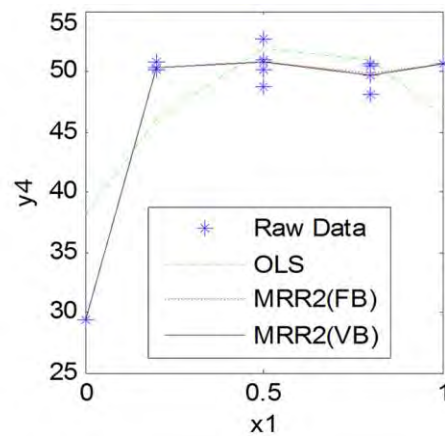


Figure 2. Plot of \hat{y}_4 vs. x_1

Figure 3 compares the plots of \hat{y}_1 versus x_1 for OLS, $MRR2_{(FB)}$, and $MRR2_{(VB)}$, when $x_2 = 0$ (left), $x_2 = 0.5$ (center), and $x_2 = 1$ (right).

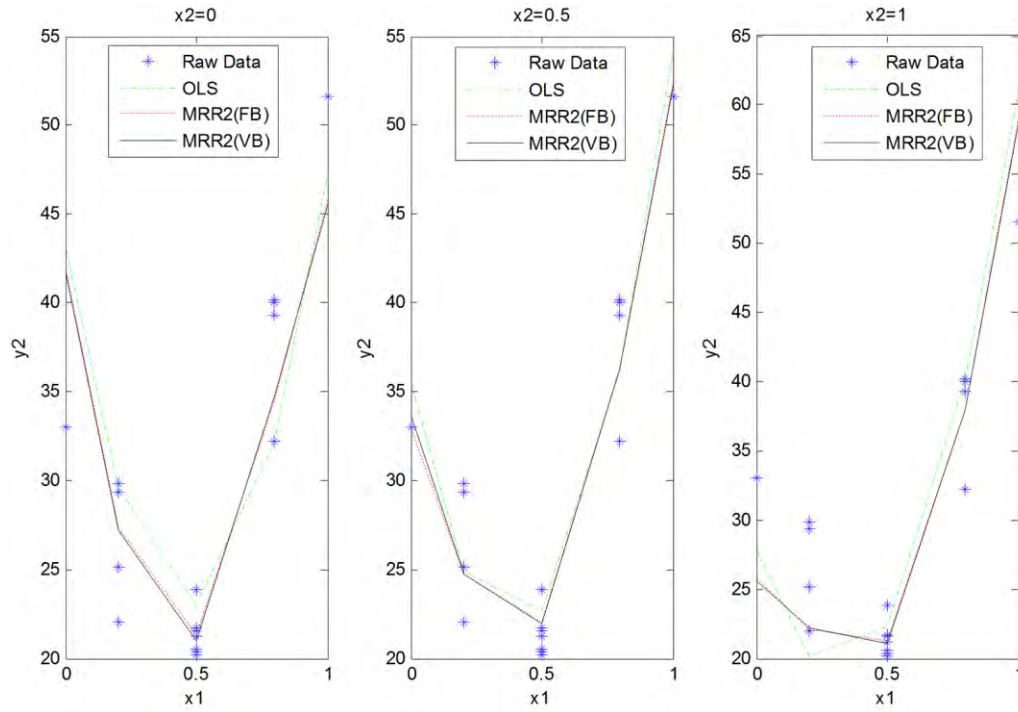


Figure 3. Plots of \hat{y}_2 versus x_1

Figure 4 compares the plots of \hat{y}_2 versus x_1 for OLS, $MRR2_{(FB)}$, and $MRR2_{(VB)}$, for all respective values of x_2 and x_3 specified

IMPROVING PERFORMANCE STATISTICS OF MRR2

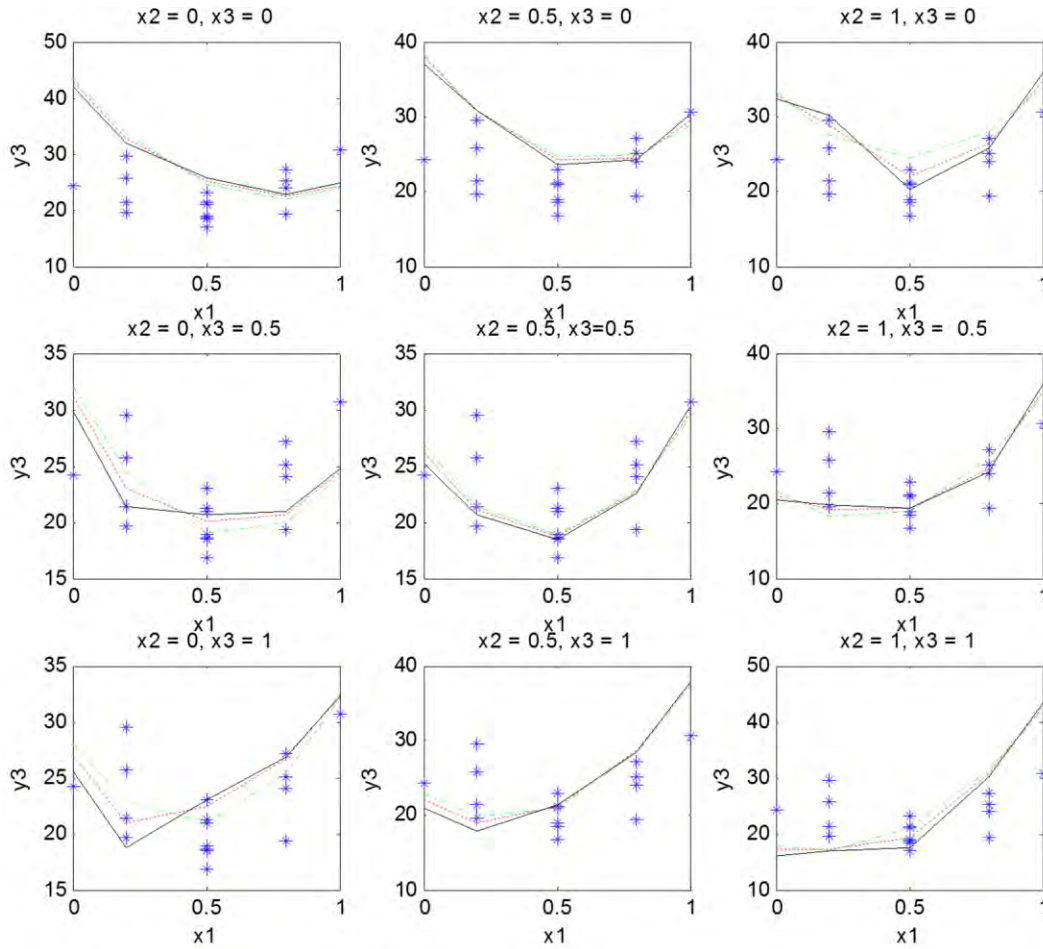


Figure 4. Plots of \hat{y}_3 versus x_1

Conclusion

One of the shortcomings of parametric regression is that the user has to specify a model that perfectly fits the data under consideration and failure to achieve this leads to highly biased estimates. Nonparametric regression is usually employed when the user is unable to specify a model for the data. However, in studies that require small-sample data such as RSM, nonparametric tends to produce fitted values that are highly variable. Semi-parametric regression such as MRR2 technique which combine parametric regression with a nonparametric technique are employed in scenarios where there is partial knowledge of the underlying model

for small-sample data. Both the nonparametric and semi-parametric methods require a parameter referred to as smoothing parameter or bandwidth which determines the smoothness of the estimates.

Regression methods for fitting data suitable for RSM were reviewed. Also reviewed are methods for selecting local bandwidth. A new methodology for deriving a function was presented. The function, herein referred to as N-squared function, was employed for generating data-driven local bandwidths and MRR2 technique utilizing local bandwidth derived from the N-squared function was applied to the multi-response problem of minced fish quality and the results of performance statistics of fitted responses was compared with the results for performance statistics for MRR2 utilizing fixed bandwidth reported in Wan (2007). The comparisons presented in Tables 4–6 show the superiority of fits from local bandwidths derived from N-squared function over fits obtained using fixed bandwidth. Indeed, these results are confirmation of statements made Wan (2007), Mays (2001a), and several other researchers regarding improvement that MRR2 and other semi-parametric methods stand to gain if performed using suitable local bandwidths.

References

- Atkeson, C. G., Moore, A. W., & Schaal, S. (1997). Locally weighted learning. *Journal of Artificial Intelligence Review*, 11, 11–73.
- Avery, M. (2010). *Literature review for local polynomial regression*. Unpublished manuscript, Department of Statistics, North Carolina State University, Raleigh, NC.
- Bickel, P. J., & Li, B. (2007). Local polynomial regression on unknown manifolds. *IMS Lecture Notes Monograph Series*, 54, 177–186.
- Del Castillo, E. (2007). *Process optimization: A statistical approach*. New York: Springer International Series in Operations Research and Management Science.
- DiNardo, J., & Tobias, J. L. (2001). Nonparametric density and regression estimation. *The Journal of Economic Perspectives*, 15(4), 11–28.
- Einsporn, R., & Birch, J. B. (1993). *Model robust regression: Using non-parametric regression to improve parametric regression analysis*. (Technical Report, 93–5). Blacksburg, VA: Department of Statistics, Virginia Polytechnic Institute & State University.

IMPROVING PERFORMANCE STATISTICS OF MRR2

- Fan J., & Gibjels I. (1992). Variable bandwidth and local linear regression smoothers. *The Annals of Statistics*, 20(4), 2008–2036.
- Galdo, J., Smith, J. & Black, D.I. (2008). Bandwidth selection and the estimation of treatment effects with unbalanced data. *Annals of Economics and Statistics*, 91-92, 189-216.
- Gerard, P. D. & Schucany, W. R. (1999). Local bandwidth selection for kernel estimation of population densities with line transect sampling. *Biometrics*, 55(3), 769-773.
- Härdle, W. (1994). *Applied nonparametric regression* (Econometric Society Monographs #19). Cambridge, England: Cambridge University Press.
- Hens, N. (2005). *Non- and semiparametric techniques for handling missing data*. (Doctoral dissertation). Diepenbeek, Belgium: Transnationale Universiteit Limburgs.
- Hernández-Lobato, J. M. (2010). *Balancing flexibility and robustness in machine learning: Semiparametric methods and sparse linear models*. (Doctoral dissertation). Madrid, Spain: University of Autonomia Department of Computer Science.
- Huang, L. S., & Fan J. (1996). Optimal bandwidth selection for local linear regression. *FSU Technical Report, N0. M.912*. Tallahassee, FL: Florida State University.
- Kohler, M., Schindler, A. & Sperlich, S. (2011). A review and comparison for bandwidth selection methods for kernel regression. *Courant Research Centre: Poverty, Equity and Growth – Discussion Papers 95*.
- Mays, J. E. (2001a). Model-robust confidence intervals and capturing nonconformity points. *Technical Report Number 01-1*, Dept. of Mathematical Sciences, Virginia Commonwealth University.
- Mays, J. E. (2001b). Small-sample model-robust confidence intervals in regression. In *Proceedings from Annual Meeting of the American Statistical Association*. Retrieved from <http://www.amstat.org/sections/SRMS/Proceedings/y2001/Proceed/00530.pdf>
- Montgomery, D. C. (1999). Experimental design for product and process design and development. *The Statistician*, 48(2), 159–177.
- Pickle, S. M. (2006). *Semiparametric techniques for response surface methodology* (Doctoral dissertation). Virginia Polytechnic Institute and State University: Blacksburg, VA.

- Racine, J. S. (2008). Nonparametric econometrics: A primer. *Journal of Foundations and Trends in Econometrics*, 3(1), 1–88.
- Raissi, S., & Farsani, R. E. (2009). Statistical process optimization through multiresponse surface methodology. *Journal of World Academy of Science, Engineering and Technology*, 3(3), 250–255.
- Schucany, W. (1995). Adaptive bandwidth choice for kernel regression. *Journal of American Statistical Association*, 90(430), 535–540.
- Starnes, B. A. (1999). *Asymptotic results for model-robust regression* (Doctoral dissertation). Virginia Polytechnic Institute and State University: Blacksburg, VA.
- Starnes, B. A., Birch, J. B., & Robinson, T. (2008, May). *A semiparametric approach to dual modeling with unreplicated data*. Paper presented at the 2008 International Conference on Interdisciplinary Mathematical and Statistical Techniques, Memphis, TN.
- Uysal, I., & Guvenir, A. (1999). An overview of regression techniques for knowledge discovery. *The Knowledge Engineering Review*, 14(4), 319–340.
- Wan, W. (2007). *Semiparametric techniques for multiresponse optimization* (Doctoral dissertation). Virginia Polytechnic Institute and State University: Blacksburg, VA.
- Wu, C. F. J., & Hamada, M. (2000). *Experiments: planning, analysis and parameter design optimization*. New York: John Wiley and Sons.
- Zheng, Q., Kulasekera, K. B. & Gallagher, C. (2010). Local adaptive smoothing in kernel regression estimation. *Statistics & Probability Letters*, 80(7-8), 540-547.

Statistical Software Applications and Review: Fitting Stereotype Logistic Regression Models for Ordinal Response Variables in Educational Research (Stata)

Xing Liu

Eastern Connecticut State University
Willimantic, CT

The stereotype logistic (SL) model is an alternative to the proportional odds (PO) model for ordinal response variables when the proportional odds assumption is violated. This model seems to be underutilized. One major reason is the constraint of current statistical software packages. Statistical Package for the Social Sciences (SPSS) cannot perform the SL regression analysis, and SAS does not have the procedure developed to directly estimate the model. The purpose of this article was to illustrate the stereotype logistic (SL) regression model, and apply it to estimate mathematics proficiency level of high school students using Stata. In addition, it compared the results of fitting the PO model and the SL model. Data from the High School Longitudinal Study of 2009 (HSL: 2009) (Ingels, et al., 2011) were used for the ordinal regression analyses.

Keywords: Stereotype logistic models, Proportional Odds models, ordinal logistic regression, ordinal response variables, Stata

Introduction

Three types of logistic regression models are well-known for analyzing the ordinal response variable, including the proportional odds (PO) model, the continuation ratio (CR) model, and the adjacent categories (AC) logistic regression model. Among them, the PO model is the most commonly used (Agresti, 2002, 2007, 2010; Armstrong & Sloan, 1989; Clogg, & Shihadeh, 1994; Hilbe, 2009; Liu, 2009; Long, 1997, Long & Freese, 2006; McCullagh, 1980; McCullagh & Nelder, 1989; O'Connell, 2000, 2006; O'Connell & Liu, 2011; Powers & Xie, 2000).

Dr. Liu is an Associate Professor in the Department of Education. Email him at liux@easternct.edu.

The PO model assumes that the underlying binary models, which dichotomize the ordinal response variable, have the same coefficients. In other words, the logit coefficients for each predictor are the same across the ordinal categories. This is called the parallel lines or the proportional odds (PO) assumption. However, the PO assumption is often violated. To deal with this issue, the partial proportional odds (PPO) model or the generalized ordinal logit model (Fu, 1998; Liu & Koirala, 2012; Peterson & Harrell, 1990; Williams, 2006) can be used. An alternative option is the stereotype logistic (SL) model, which was first developed by Anderson (1984), and later introduced by Greenland (1994), and Long and Freese (2006). The SL model is an extension of both the multinomial logistic regression model and the PO model. First, the SL model is like the multinomial logistic model since they both estimate the odds of being at a particular category compared to the baseline category. Second, similar to the PO model, the SL model estimates the ordinal response variable rather than the nominal outcome variable, given a set of predictors. However, the SL model does not assume the PO assumption, and allows the effect of each predictor to vary across the ordinal categories.

Although the theory of the SL model has existed, this model seemed to be underutilized: the illustration and application of this model were rare. One major reason is the restriction of current statistical software packages. SPSS cannot perform the SL regression analysis, and SAS does not have the procedure developed to directly estimate the SL model. Both Anderson (1984) and Greenland (1994) used GAUSS to fit the SL model but no programming information was provided. Agresti (2010) recently discussed this model using the results of the two examples directly from Anderson (1984). Kuss (2006) pioneered the use the PROC NLMIXED procedure in SAS to estimate the SL model although it does not deal with any random effects in the example. Researchers need to specify the starting values, and the model equations, and the probabilities in the syntax, which is complicated and error-prone for novice SAS users. Therefore, it is critical to help researchers to familiarize with this model and clarify the confusion so that they are able to apply it correctly in practice.

To fill this gap, the purpose of this study was to illustrate the use of the stereotype logistic (SL) regression with Stata, and compare the results of fitting the PO model and the SL model. This article is an extension of previous research on various ordinal logistic regression models (Liu, 2009; Liu, O'Connell & Koirala, 2011; Liu & Koirala, 2012; O'Connell & Liu, 2010). For demonstration purposes, the empirical data from the High School Longitudinal Study of 2009 (HSLs: 2009) (Ingels, et al., 2011) were used to conduct the ordinal regression analyses.

Theoretical Framework

The Proportional Odds Model

An ordinal logistic regression model is a generalization of a binary logistic regression model, when the outcome variable has more than two ordinal levels. It estimates the cumulative odds and the probability of an observation being at or below a specific outcome level, conditional on a collection of explanatory variables. In Stata, the ordinal logistic regression model assumes that the outcome variable is a latent variable, which is expressed in logit form as follows

$$\ln(Y'_i) = \text{logit}[\pi(x)] = \ln\left(\frac{\pi_j(x)}{1 - \pi_j(x)}\right) = \alpha_j + (-\beta_1 X_1 - \beta_2 X_2 - \dots - \beta_p X_p) \quad (1)$$

where $\pi_j(x) = \pi(Y \leq j | x_1, x_2, \dots, x_p)$, which is the probability of being at or below category j , given a set of predictors $j = 1, 2, \dots, J-1$. α_j are the cut points, and $\beta_1, \beta_2, \dots, \beta_p$ are logit coefficients. This is also known as the proportional odds (PO) model because the odds ratio of any predictor is assumed to be constant across all categories. Therefore, for each predictor, there is only one logit coefficient across all the comparisons, i.e., at or below a certain category versus above that category. The Brant test is used to assess the proportional odds assumption (Brant, 1990). To estimate the $\ln(\text{odds})$ of being at or below the j^{th} category, the PO model can be rewritten as

$$\begin{aligned} \text{logit}[\pi(Y \leq j | x_1, x_2, \dots, x_p)] &= \ln\left(\frac{\pi(Y \leq j | x_1, x_2, \dots, x_p)}{\pi(Y > j | x_1, x_2, \dots, x_p)}\right) \\ &= \alpha_j + (-\beta_1 X_1 - \beta_2 X_2 - \dots - \beta_p X_p) \end{aligned} \quad (2)$$

Thus, this model predicts cumulative logits across $J-1$ response categories. By transforming the cumulative logits, we can obtain the estimated cumulative odds as well as the cumulative probabilities being at or below the j^{th} category.

Researchers may see different forms of the ordinal logistic regression model in literature since different software packages may employ different parameterizations when estimating logit coefficients (Liu, 2009). For example, SPSS uses the same form as that in Stata. However, SAS uses a different form where a positive sign is placed before logit coefficients.

The Multinomial Logistic Model

The multinomial logistic regression model is also an extension of the binary logistic regression model when the outcome variable is nominal and has more than two categories. It estimates the odds of being at any category compared to being at the baseline category, also called the comparison category. It can be treated as a combination of a series of binary logistic regression models with a particular category = 1, and the base category = 0. When there are J categories, it estimates $J-1$ binary logistic regression models simultaneously. This model can be expressed as follows:

$$\ln \left(\frac{\pi(Y = j | x_1, x_2, \dots, x_p)}{\pi(Y = J | x_1, x_2, \dots, x_p)} \right) = \alpha_j + \beta_{j1}X_1 + \beta_{j2}X_2 + \dots + \beta_{jp}X_p \quad (3)$$

where $j = 1, 2, \dots, J-1$; J is the base category, which can be any category but is generally the highest one; α_j are the intercepts, and $\beta_{j1}, \beta_{j2}, \dots, \beta_{jp}$ are logit coefficients. Since the model includes $J-1$ comparisons, it estimates $J-1$ logit coefficients for each predictor.

The Stereotype Logistic Model

Anderson's SL model (1984) can be written in the following form

$$\begin{aligned} \text{logit}[\pi(j, J)] &= \ln \left(\frac{\pi(Y = j | x_1, x_2, \dots, x_p)}{\pi(Y = J | x_1, x_2, \dots, x_p)} \right) \\ &= \alpha_j - \phi_j (\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p) \end{aligned} \quad (4)$$

where $j = 1, 2, \dots, J-1$; J is the baseline or reference category, which is the last category here, but can be the first category or any of the other categories decided by the researcher; Y is the ordinal response variable with categories from j to J ; α_j are the intercepts; $\beta_1, \beta_2, \dots, \beta_p$ are logit coefficients for the predictors, X_1, X_2, \dots, X_p , respectively, and ϕ_j are the constraints which are used to ensure the outcome variable is ordinal if the following condition is satisfied.

$$1 = \phi_1 > \phi_2 > \phi_3 > \dots > \phi_{J-1} > \phi_J = 0 \quad (5)$$

The first constraint, ϕ_1 is set to be 1, and the last one, ϕ_J is equal to 0 so that the estimated SL model can be identified. If any two pairs of the constraints are the same, then these two categories are indistinguishable, thus can be collapsed into one. For example, if $\phi_3 = \phi_4$, these two categories (categories 3 and 4) can be grouped together. The ordinality of the constraints can be tested in the model so that researchers can decide whether any categories need to be merged or re-ordered.

To calculate the odds of being in a category j versus a category m , we just need to take the exponential of $[(\alpha_j - \alpha_m) - (\phi_j - \phi_m)\beta]$. When the category m becomes the baseline category J , we just need to substitute it into the equation. Since $\phi_J = 0$, we get $[(\alpha_j - 0) - (\phi_j - 0)\beta] = \alpha_j - \phi_j\beta$. By exponentiating $(-\phi_j\beta)$, we get the odds of being in a category j versus the baseline category J for a unit change in a predictor.

The equation (4) is the forms for Anderson's one-dimension SL model, which was generally referred to as the SL model in literature. Anderson (1984) also argued that an ordinal response variable could be more than one dimension, and therefore proposed the multidimensional SL model. If the ordinal outcome variable has J categories, the maximum dimensions would be $J-1$. The multidimensional SL model with $J-1$ dimensions is actually equal to the multinomial logistic regression model. In this article, we only focus on the one-dimension SL model for the simplicity of model building and interpretation.

Lunt (2001) considered the SL model as the constrained multinomial logistic model, and developed the Stata *soreg* program before the official Stata *slogit* program was implemented. Compared with the multinomial logistic regression model in the equation (3), the left side of the logit link function for the SL model in the equation (4) looks the same, since both the SL model and the multinomial model estimates the odds of being in a particular category versus the baseline category. Examining the systematic component (linear predictors) in both models, it is obvious that the logit coefficients, β_j in the multinomial logistic model corresponds to $(-\phi_j(\beta))$ in the SL model. When there are J categories of the outcome variable and p predictors, we need to estimate $(J-1) + (J-1)\times p$ parameters in the multinomial logistic model, which also equals $(J-1)\times(1+p)$. In the SL model, we estimate $[(J-1) + (J-2) + p] = (2J-3+p)$ parameters since ϕ_1 and ϕ_J are constrained to be 1 and 0, respectively. Therefore, less parameters are estimated in the SL model than in the multinomial logistic model, and the former model is more parsimonious.

Methodology

Sample

Similar to the previous Education Longitudinal Study of 2002 (ELS: 2002), the HSLS: 2009 study, conducted by the NCES, was the latest series of longitudinal study in secondary schools. This study surveyed high school students, parents, teachers, school counselors and administrators, and assessed 9th graders' algebraic skills and reasoning. It was designed to keep track of high school students from grade nine to postsecondary school education and their choice of future careers. In the 2009 base year data, 21,444 high school students, from a national sample of 944 schools, participated in the study. Students were asked to provide information regarding basic demographics, school and home experience, such as math and science activities, coursework, and time spent on different activities, mathematics and science attitude, mathematics and science self-efficacy, their feelings about math and science teacher, and future educational and life plans after secondary schools. The ordinal outcome variable is students' mathematics proficiency, and the predictors are students' math identity (MTHID), mathematics self-efficacy (MTHEFF), school belonging (SCHBEL), and school engagement (SCHENG).

The outcome variable, students' mathematics proficiency levels in high schools, was ordinal with five levels (1 = students can answer questions in algebraic expressions; 2 = students can answer questions and solve problems for multiplicative and proportional situations; 3 = students can understand algebraic equivalents and solve problems; 4 = students can understand systems of linear equations and solve problems; 5 = students can understand linear functions, find and use slopes and intercepts of lines, and can use functional notation) (Ingels, et al., 2011). In addition, those students who failed to pass through level 1 were assigned to level 0. Table 1 provides the frequency of six mathematics proficiency levels (from 0 to 5).

Table 1. Proficiency categories and frequencies (proportions) for the study sample, HSLS: 2009 (Ingels, et al., 2011) base year

Category	Description	Frequency
0	Did not pass level 1	2263 (10.6%)
1	Algebraic expressions	4933 (23%)
2	Multiplicative and proportional thinking	5495 (25.6%)
3	Algebraic equivalents	5761 (26.9%)
4	Systems of equations	2396 (11.2%)
5	Linear functions.	596 (2.8%)

STEREOTYPE LOGISTIC REGRESSION MODELS

Data Analysis

First, the PO model was used for the preliminary analysis with the Stata *ologit* command, and the proportional odds assumption was examined using the Brant test. Then the SL model with a single explanatory variable was fitted using the Stata *slogit* command. Finally the full-model with all four explanatory variables was fitted. Model fit statistics for both the PO model and the SL model were provided by the Stata *SPost* package (Long & Freese, 2006). The results for both models were interpreted and compared. Following the suggestion by Hardin and Hilbe (2007) and Hilbe (2009), Stata AIC and BIC statistics were used for the comparison of model fit.

Results

The Proportional Odds Model with Four Explanatory Variables

A PO model with all four predictor variables was fitted first, since it is the most commonly used model for ordinal response variables. The Stata *ologit* command with the logit function was used for model fitting. Figure 1 displays the Stata output for the PO model.

```
. ologit Mathprof MTHID MTHEFF SCHBEL SCHENG

Iteration 0:  log likelihood = -28870.574
Iteration 1:  log likelihood = -27172.197
Iteration 2:  log likelihood = -27148.796
Iteration 3:  log likelihood = -27148.754

Ordered logistic regression
LR chi2(4)      =    3443.64
Prob > chi2     =    0.0000
Log likelihood = -27148.754

Number of obs   =    17848
Pseudo R2      =    0.0596

-----+-----
Mathprof |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
MTHID |      .5951592   .0171736    34.66   0.000   .5614995   .628819
MTHEFF |      .1884678   .0172434    10.93   0.000   .1546714   .2222643
SCHBEL |      .0893507   .014859    6.01    0.000   .0602276   .1184738
SCHENG |      .224087    .0149587    14.98   0.000   .1947684   .2534055
-----+-----
/cut1 |     -2.560906   .0278366          -2.615465   -2.506348
/cut2 |     -.8916257   .0172818          -.9254974   -.857754
/cut3 |      .3439423   .0160976           .3123915   .375493
/cut4 |      1.967861   .022283          1.924187   2.011535
/cut5 |      3.788231   .0440731          3.701849   3.874612
-----+-----
```

Figure 1. Stata Proportional Odds model with four predictor variables

The log likelihood ratio Chi-Square test, $LR \chi^2_{(4)} = 3443.64$, $p < .001$, indicating that the full model with four predictor provided a better fit than the null model with no independent variables.

The fit statistics for the model were as follows. The likelihood ratio $R^2_L = .060$, Cox-Snell $R^2 = .175$, Nagelkerke $R^2 = .183$, AIC = 3.043, AIC used by Stata = 54315.508, and BIC = -120340.00. A summary of more detailed fit statistics is provided in Figure 2. Both AIC and AIC used by Stata in the PO model were used as the base for future model comparisons.

The logit effects of all four predictors on the ordinal response variable, mathematics proficiency were significant. The estimated logit regression coefficient for math identity (*mthid*), $\beta = .595$, $z = 34.66$, $p < .001$; the logit coefficient for mathematics self-efficacy (*mtheff*), $\beta = .188$, $z = 10.93$, $p < .001$; the coefficient for school belonging (*schbel*), $\beta = .089$, $z = 6.01$, $p < .001$, and finally, for school engagement (*scheng*), $\beta = .224$, $z = 14.98$, $p < .001$. All four predictors were positively associated with the log odds of being beyond a proficiency level. In terms of odds ratio (OR), the odds of being beyond a proficiency level were 1.813 times greater with a one unit increase in higher level of math identity, and 1.207 times greater with one unit increase in students' mathematics self-efficacy. In addition, students who had higher level of school belong and school engagement were more likely to be associated with higher level of mathematics proficiency (ORs = 1.093 and 1.251 for *schbel* and *scheng*, respectively).

```
. fitstat
```

Measures of Fit for ologit of Mathprof			
Log-Lik Intercept Only:	-28870.574	Log-Lik Full Model:	-27148.754
D(17839):	54297.508	LR(4):	3443.639
		Prob > LR:	0.000
McFadden's R2:	0.060	McFadden's Adj R2:	0.059
ML (Cox-Snell) R2:	0.175	Cragg-Uhler(Nagelkerke) R2:	0.183
McKelvey & Zavoina's R2:	0.174		
Variance of y*:	3.983	Variance of error:	3.290
Count R2:	0.331	Adj Count R2:	0.068
AIC:	3.043	AIC*n:	54315.508
BIC:	-120340.000	BIC':	-3404.481
BIC used by Stata:	54385.615	AIC used by Stata:	54315.508

Figure 2. Fit statistics for the PO model

Brant Test of the Proportional Odds Assumption

The Brant test of the PO assumption was examined using the *brant* command of the Stata *SPost* (Long & Freese, 2006) package. Stata Brant test provided results of a series of separate binary logistic regression across different category comparisons, univariate brant test result for each predictor, and the omnibus test for the overall model. Table 2 provides five ($j-1 = 5$) associated binary logistic regression models for the full PO model, where each split compares $Y > \text{cat. } j$ to $Y \leq \text{cat. } j$, since data were dichotomized according to probability comparisons. Among the logit coefficient of all four variables across five logistic regression models, only the effect of school belonging was similar across these models. The coefficient of math identify was similar across the first three models, but it started to increase from the model 4 to 5; the logit coefficient in model 5 was almost the double of that in model 1. Although the coefficients of school engagement looked similar across the models, those for the models 1 and 5 were the largest. The coefficients of mathematics self-efficacy 1 were stable across the first four logistic regression models, but it increased abruptly in model 5.

To test the PO assumptions, the Brant test provided the results for the overall model and each predictor. Table 3 presents χ^2 tests and p values for the full PO model and separate variables. The omnibus Brant test for the full model, $\chi^2_{16} = 178.52$, $p < .001$, indicating that the proportional odds assumption for the full model was violated. To identify which predictor variables violated the assumption, separate Brant tests were examined for each predictor variable. The results revealed that the univariate Brant tests for the PO assumption were upheld for using computers for fun, and using computers to learn on own. On the other hand, the Brant test was violated for using computers for school work.

Table 2. A Series ($j-1$) of Associated Binary Logistic Regression models for the full PO model, where each split compares $Y > \text{cat. } j$ to $Y \leq \text{cat. } j$

	Y > 0	Y > 1	Y > 2	Y > 3	Y > 4	Brant Test P Value
Variable	Logit (b)	Logit (b)	Logit (b)	Logit (b)	Logit (b)	
Constant	2.529	0.859	-0.374	-2.103	-4.310	
<i>mthid</i>	0.528	0.508	0.583	0.804	1.082	.000**
<i>mtheff</i>	0.150	0.184	0.212	0.179	0.332	.033*
<i>schbel</i>	0.088	0.095	0.098	0.074	0.015	.577
<i>scheng</i>	0.286	0.206	0.226	0.235	0.290	.018*

Note: * $p < .05$; ** $p < .01$

Table 3. Brant tests of the PO assumption for each predictor and the overall model

Variable	Test	P Value
<i>mthid</i>	$\chi^2_4 = 101.01$.000**
<i>mtheff</i>	$\chi^2_4 = 10.48$.033*
<i>schbel</i>	$\chi^2_4 = 2.88$.577
<i>scheng</i>	$\chi^2_4 = 11.91$.018*
All (Full-model)	$\chi^2_{16} = 178.52$.000**

Note: * $p < .05$; ** $p < .01$

The Stereotype Logistic Regression Model with a Single Explanatory Variable

Stereotype logistic regression models were fitted since they released the PO assumption and allowed the logit coefficients to vary across the ordinal categories. For comparison purposes, model fitting process included both a single variable model and the full model with all four predictor variables. Figure 3 presents the Stata output for the single predictor SL model.

The Wald Chi-Square test with 1 degree of freedom, Wald $\chi^2_{(1)} = 1044.37$, $p < .001$, indicating that the logit coefficient of the predictor, math identity was statistically different from 0. Since no R^2 statistics were calculated, only the AIC and BIC statistics were reported. The AIC statistic was 3.072, and the AIC used by Stata was 64996.139. BIC was -145654.148 , and the corresponding BIC used by Stat was 65075.737.

The estimated logit coefficient, $\beta = 2.116$, $z = 32.32$, $p < .001$, indicating that students' math identity had a significant relationship with mathematics proficiency.

The SL model estimates the logit odds of being in a category relative to the baseline category. Substituting the value of the coefficient into the formula (4)

$$\ln \left(\frac{\pi(Y = j | x_1)}{\pi(Y = J | x_1)} \right) = \alpha_j - \phi_j(\beta_1 X_1),$$

we calculated

$$\text{logit}[\pi(j, J | \text{mathid})] = \alpha_j - \phi_j * 2.116(\text{mathid}).$$

STEREOTYPE LOGISTIC REGRESSION MODELS

```
. slogit Mathprof MTHID, dim(1)
```

```
Iteration 0:  log likelihood = -34026.719   (not concave)
Iteration 1:  log likelihood = -32684.99   (not concave)
Iteration 2:  log likelihood = -32551.807
Iteration 3:  log likelihood = -32499.066
Iteration 4:  log likelihood = -32488.31
Iteration 5:  log likelihood = -32488.069
Iteration 6:  log likelihood = -32488.069
```

```
Stereotype logistic regression      Number of obs   =      21159
Wald chi2(1)                       =      1044.37
Log likelihood = -32488.069         Prob > chi2      =      0.0000
```

```
( 1) [phi1_1]_cons = 1
```

Mathprof	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
MTHID	2.115636	.0654657	32.32	0.000	1.987326 2.243947
/phi1_1	1
/phi1_2	.8808938	.012357	71.29	0.000	.8566745 .905113
/phi1_3	.750083	.0126924	59.10	0.000	.7252063 .7749598
/phi1_4	.5680588	.0152125	37.34	0.000	.5382429 .5978746
/phi1_5	.3290306	.0223091	14.75	0.000	.2853056 .3727555
/phi1_6	0 (base outcome)				
/theta1	2.039036	.0797773	25.56	0.000	1.882676 2.195397
/theta2	2.951371	.0773612	38.15	0.000	2.799746 3.102997
/theta3	3.11343	.0771105	40.38	0.000	2.962297 3.264564
/theta4	3.119226	.0771187	40.45	0.000	2.968076 3.270376
/theta5	2.001495	.0800912	24.99	0.000	1.844519 2.158471
/theta6	0 (base outcome)				

```
(Mathprof=5 is the base outcome)
. fitstat
```

```
Measures of Fit for slogit of Mathprof
```

Log-Lik Full Model:	-32488.069	D(21148):	64976.139
Wald X2(1):	1044.369	Prob > X2:	0.000
AIC:	3.072	AIC*n:	64998.139
BIC:	-145654.148		
BIC used by Stata:	65075.737	AIC used by Stata:	64996.139

Figure 3. The Stereotype Logistic Regression model: Single Predictor, Math Identity

Recall that ϕ_j is a list of ordinality constraints with the first constraint = 1 and the last one = 0, and it satisfies the condition $1 = \phi_1 > \phi_2 > \phi_3 > \dots \phi_{J-1} > \phi_J = 0$. The estimated ϕ_j s in the model were as follows: 1, .881, .750, .568, .329, and 0, which were used to ensure the ordering of the mathematics proficiency level.

The odds ratio of being in a category j versus the baseline category J was obtained by taking the exponential of $[(\alpha_j - \alpha_J) - (\phi_j - \phi_J)\beta]$. Since the baseline category J was the mathematics proficiency level 5 in the model, the estimated α_J and ϕ_J were 0, and then the equation could be simplified to be $(\alpha_j - \phi_j\beta)$. By exponentiating $(-\phi_j\beta)$, we got the odds ratio of being a category j versus the baseline J for a unit change in a predictor variable. In this model, the odds ratio of being in

mathematics proficiency level 0 compared to being in level 5, $OR(0,5) = e^{(-1*2.116)} = e^{(-2.116)} = .121$. This indicated that for a unit increase in math identity the odds of being in mathematics proficiency level 0 compared to being the baseline category 5 decreased by a factor of .121. In other words, students were more likely to be in the highest proficiency level 5 rather than being in level 0 when students had higher level of math identity.

Since $\phi_2 = .881$, the odds ratio of being in mathematics proficiency level 1 compared to being in level 5, $OR(1,5) = e^{(-.881*2.116)} = e^{(-1.864)} = .155$. Since ϕ_3 , ϕ_4 , and ϕ_5 were .750, .568, and .329, respectively, the odds ratio of being in the other proficiency levels compared to being in the baseline level were calculated in the same way. $OR(2,5)$, $OR(3,5)$ and $OR(4,5)$ were .205, .301, and .498 respectively.

The odds of being in the baseline category J , relative to a particular category j , is the inverse of the odds of being in that category versus the baseline category. To estimate the odds of being in the baseline category relative to a particular category, we just need to change the signs before the cutpoints and the estimated logits in the equation (6). The modified logit equation, $\text{logit}[\pi(J, j | mthid)] = -\alpha_j + \phi_j \times 2.116(mthid)$. By exponentiating $(\phi_j\beta)$, we get the odds ratio of being in the baseline category J versus any other category for a one unit change in a predictor variable.

$OR(5,0) = e^{(1*2.116)} = 8.295$, indicating that the odds of being in the proficiency level 5 relative to the level 0 were 8.295 times greater with one unit increase in math identity. The odds ratio of being in the baseline level 5 compared to being in level 1, $OR(5,1) = e^{(.881*2.116)} = e^{(1.864)} = 6.449$. The ORs of being the baseline category versus the other three categories were computed in the same way, and they were 4.889, 3.326, and 2.006, respectively.

The Full Stereotype Logistic Regression Model with Four Predictor Variables

Next, the full SL model with all four predictor variables was fitted. Figure 4 and Table 4 provide the results of the full SL model. The Wald Chi-Square test, Wald $\chi^2_{(4)} = 1145.98$, $p < .001$, indicating that the full model provides a better fit than the null model. The AIC statistic was 3.034, and the AIC used by Stata was 54156.535. BIC was -120458.025, and the corresponding BIC used by Stata was 54257.800.

STEREOTYPE LOGISTIC REGRESSION MODELS

```
. slogit Mathprof MTHID MTHEFF SCHBEL SCHENG, dim(1) nolog
```

Stereotype logistic regression Number of obs = 17848
 Wald chi2(4) = 1145.98
 Log likelihood = -27065.267 Prob > chi2 = 0.0000

(1) [phi1_1]_cons = 1

Mathprof	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
MTHID	1.654606	.0622347	26.59	0.000	1.532628	1.776584
MTHEFF	.530026	.0485596	10.91	0.000	.4348509	.6252012
SCHBEL	.2235734	.0400664	5.58	0.000	.1450447	.302102
SCHENG	.6247203	.0436208	14.32	0.000	.5392251	.7102154
/phi1_1	1
/phi1_2	.8486203	.0129488	65.54	0.000	.8232411	.8739996
/phi1_3	.7215881	.0131181	55.01	0.000	.6958772	.747299
/phi1_4	.5239136	.0154778	33.85	0.000	.4935777	.5542495
/phi1_5	.298745	.0214996	13.90	0.000	.2566065	.3408835
/phi1_6	0 (base outcome)					
/theta1	1.778747	.084659	21.01	0.000	1.612819	1.944676
/theta2	2.858481	.0808379	35.36	0.000	2.700042	3.016921
/theta3	3.084861	.0804567	38.34	0.000	2.927169	3.242553
/theta4	3.104532	.0804757	38.58	0.000	2.946803	3.262262
/theta5	2.000653	.0836264	23.92	0.000	1.836748	2.164558
/theta6	0 (base outcome)					

(Mathprof=5 is the base outcome)

```
. fitstat
```

Measures of Fit for slogit of Mathprof

Log-Lik Full Model:	-27065.267	D(17834):	54130.535
Wald X2(4):	1145.978	Prob > X2:	0.000
AIC:	3.034	AIC*n:	54158.535
BIC:	-120458.025		
BIC used by Stata:	54257.800	AIC used by Stata:	54156.535

Figure 4. The Stereotype Logistic Regression model: Full Model

Compared with the single-variable SL model (see Table 4), both AIC and AIC by Stata indicated that the full-model fitted the data better (3.034 and 54156.535, respectively for the full-model vs. 3.072 and 64996.139, respectively for the single model). This result was also supported by the model comparison using the BIC and BIC by Stata.

Compared with the PO model (AIC = 3.043, and AIC used by Stata = 54315.508), the full SL model also had a better fit, which indicated that the SL model was a better choice when the proportional odds assumption was untenable in the PO model.

The logit effects of all four predictor variables on mathematics proficiency were significant. Similar to that in the single variable SL model, the estimated logit regression coefficient for math identify (*mthid*), $\beta = 1.655$, $z = 26.59$, $p < .001$; the

logit coefficient for mathematics self-efficacy (*mtheff*), $\beta = .530$, $z = 10.91$, $p < .001$; the logit coefficient for school belonging (*schbel*), $\beta = .224$, $z = 5.58$, $p < .001$; and finally, for school engagement (*scheng*), $\beta = .625$, $z = 14.32$, $p < .001$. These logit coefficients compared the probabilities of being in the baseline category versus the lowest category. The all four predictor variables were positively associated with the odds of being in the baseline level 5 compared to the level 0. In terms of odds ratio (OR), the odds of being in the baseline proficiency level 5 versus the level 0 increased by a factor of 5.231 with a one unit increase in math identity; they increased by a factor of 1.699 for a one unit increase in mathematics self-efficacy; they increased by a factor of 1.251 for school belonging; and finally they increased by a factor of 1.868 for school engagement, holding the effects of the other variables constant.

Table 4. Results of the Single-Variable SL Model and the Full SL Model

Variable	Single-Variable Model		Full Model	
	b (se(b))	OR	b (se(b))	OR
α_1	2.039 (.080)		1.779 (.085)	
α_2	2.951 (.077)		2.858 (.081)	
α_3	3.113 (.077)		3.084 (.080)	
α_4	3.119 (.077)		3.104 (.080)	
α_5	2.001 (.080)		2.000 (.084)	
α_6	0 (base)		0 (base)	
ϕ_1	1		1	
ϕ_2	.881 (.012)		.849 (.013)	
ϕ_3	.750 (.013)		.722 (.013)	
ϕ_4	.568 (.015)		.524 (.015)	
ϕ_5	.329 (.022)		.299 (.022)	
ϕ_6	0		0	
<i>MTHID</i>	2.116 (.065)**	8.295	1.655(.062)**	5.231
<i>MTHEFF</i>			.530 (.049)**	1.699
<i>SCHBEL</i>			.224 (.040)**	1.251
<i>SCHENG</i>			.625 (.044)**	1.868
Model Fit	$\chi^2_1 = 1044.37^{**}$		$\chi^2_4 = 1145.98^{**}$	
AIC	3.072		3.034	
AIC by Stata	64996.139		54156.535	

STEREOTYPE LOGISTIC REGRESSION MODELS

Just as the single-predictor SL model, by exponentiating $(\phi_j\beta)$ for each of the four predictor variables, we obtain the odds of being in the baseline category J versus any other category. Table 5 shows the odds comparing the baseline category and the other categories for all four predictor variables.

Table 5. Odds ratios for all four predictor variables across five comparisons ($Y = J$ vs. $Y = j$)

Category Comparison	Y=5 vs. Y=0	Y=5 vs. Y=1	Y=5 vs. Y=2	Y=5 vs. Y=3	Y=5 vs. Y=4
Variables	OR	OR	OR	OR	OR
<i>mathid</i>	5.231	4.072	3.3	2.379	1.639
<i>mtheff</i>	1.699	1.568	1.466	1.32	1.172
<i>schbel</i>	1.251	1.209	1.175	1.124	1.069
<i>scheng</i>	1.868	1.699	1.57	1.387	1.205

Conclusions

The use of stereotype logistic models was used to estimate ordinal mathematics proficiency using Stata when the proportional odds assumption is not upheld. The PO model with Stata was fitted first for the preliminary analysis, and then the proportional odds assumption was tested. The results of the Brant test indicated that the proportional odds assumption was violated. We then fitted the SL models starting from a single-variable model to the full model with four predictor variables. Finally, results of the PO model and the SL model were interpreted and compared.

Compared to the PO model, it is found that the SL model is a better option when the proportional odds assumption is untenable. The SL model not only relaxes the PO assumption but also ensures the ordinal information of the categorical variable by putting an ordinality constraint on the estimated coefficients.

It should be noted that the interpretations of the odds ratios in the PO model and the SL model are different. While the PO models estimate the cumulative odds of being at or below a particular category relative to above that category, the SL models estimate the odds of being at a category relative to the baseline category. In addition, to calculate the odds ratios in the SL model, we need to take both the ordinality constraints and the logit coefficients into consideration. In other words, we need to take the exponential of the product of the ordinality constraints and the coefficients.

Alternative to the SL model, another option dealing with the violation of the proportional odds assumption is the partial proportional odds (PPO) model or the generalized ordinal logit model. Interested researchers may refer to Peterson and Harrell (1990) for theories of the PPO model, Fu (1998), Liu and Koirala (2012), and Williams (2006) for the illustration of both models using Stata, and O’Connell (2006), and Stoke, Davis and Koch (2000) for the illustration of the PPO model using SAS.

Because the SL model is not widely available in other statistical software packages, the focus was only on the illustration of the use this model in Stata. Future research will be extended to other software packages once they make the SL model available. It is our hope that researchers could familiarize with the SL model and apply it correctly in their own research.

Acknowledgements

Previous versions of this paper were presented at the Modern Modeling Methods Conference in Storrs, CT (May, 2013), the Northeastern Educational Research Association Annual Conference in Rocky Hill, CT (Oct., 2013), and the Annual Meeting of American Educational Research Association (AERA) in Philadelphia, PA (April 2014).

References

- Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New York: John Wiley and Sons.
- Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). New York: John Wiley and Sons.
- Agresti, A. (2010). *The analysis of ordinal categorical data* (2nd ed.). New York: John Wiley and Sons.
- Anderson, J. A. (1984). Regression and ordered categorical variables. *Journal of Royal Statistical Society, Series B*, 46, 1-30.
- Armstrong, B. B., & Sloan, M. (1989). Ordinal regression models for epidemiological data. *American Journal of Epidemiology*, 129(1), 191-204.
- Brant. (1990). Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics*, 46, 1171-1178.
- Clogg, C. C., & Shihadeh, E. S. (1994). *Statistical models for ordinal variables*. Thousand Oaks, CA: Sage
- Fu, V. (1998). Estimating generalized ordered logit models. *Stata Technical Bulletin*, 44, 27-30.
- Greenland, S. (1994). Alternative models for ordinal logistic regression. *Statistics in Medicine*, 13(16), 1665-1677.
- Hardin, J. W., & Hilbe, J. M. (2007). *Generalized linear models and extensions* (2nd ed.). Texas: Stata Press.
- Hilbe, J. M. (2009). *Logistic regression models*. Boca Raton, FL: Chapman & Hall/CRC.
- Ingels, S. J., Dalton, B., Holder, T. E., Lauff, E., & Burn, L. J. (2011). *The High School Longitudinal Study of 2009 (HSLs:09): A first look at fall 2009 ninth-graders* (NCES 2011-327). U.S. Department of Education. Washington DC: National Center for Education Statistics.
- Kuss, O. (2006). On the estimation of the stereotype regression model. *Computational Statistics & Data Analysis*, 50, 1877-1890.
- Liu, X. (2009). Ordinal regression analysis: Fitting the proportional odds model using Stata, SAS and SPSS. *Journal of Modern Applied Statistical Methods*, 8(2), 632-645. Retrieved from <http://digitalcommons.wayne.edu/jmasm/vol8/iss2/30/>
- Liu, X., & Koirala, H. (2012). Ordinal regression analysis: Using generalized ordinal logistic regression models to estimate educational data.

Journal of Modern Applied Statistical Methods, 11(1), 242-254. Retrieved from <http://digitalcommons.wayne.edu/jmasm/vol11/iss1/21/>

Liu, X., O'Connell, A. A., & Koirala, H. (2011). Ordinal regression analysis: Predicting mathematics proficiency using the continuation ratio model. *Journal of Modern Applied Statistical Methods*, 10(2), 513-527. Retrieved from <http://digitalcommons.wayne.edu/jmasm/vol10/iss2/11/>

Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage.

Long, J. S. & Freese, J. (2006). *Regression models for categorical dependent variables using Stata* (2nd ed.). Texas: Stata Press.

Lunt, M. (2001). Stereotype ordinal regression. *Stata Technical Bulletin*, 61, 12-18.

McCullagh, P. (1980). Regression models for ordinal data (with discussion). *Journal of the Royal Statistical Society Ser. B*, 42, 109-142.

McCullagh, P. & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). London: Chapman and Hall.

O'Connell, A. A., (2000). Methods for modeling ordinal outcome variables. *Measurement and Evaluation in Counseling and Development*, 33(3), 170-193.

O'Connell, A. A. (2006). *Logistic regression models for ordinal response variables*. Thousand Oaks: SAGE.

O'Connell, A. A., & Liu, X. (2011). Model diagnostics for proportional and partial proportional odds models. *Journal of Modern Applied Statistical Methods*, 10(1), 139-175. Retrieved from <http://digitalcommons.wayne.edu/jmasm/vol10/iss1/15/>

Peterson, B., & Harrell, F. E. (1990). Partial proportional odds models for ordinal response variables. *Applied Statistics*, 39(2), 205-217.

Powers D. A., & Xie, Y. (2000). *Statistical models for categorical data analysis*. San Diego, CA: Academic Press.

Stokes, M. E., Davis, C. S., & Koch, G. G. (2000). *Categorical Data Analysis Using the SAS System*. Cary, NC: SAS Institute Inc.

Williams, R. (2006). Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *The Stata Journal*, 6(1), 58-82.

Instructions for Authors

Authors wishing to submit to *JMASM* may do so using the submission form at the journal's website, <http://digitalcommons.wayne.edu/jmasm>. Three areas are appropriate for *JMASM*:

1. Development or study of new statistical tests or procedures, or the comparison of existing statistical tests or procedures, using computer-intensive Monte Carlo, bootstrap, jackknife, or resampling methods;
2. Development or study of nonparametric, robust, permutation, exact, and approximate randomization methods; and
3. Applications of computer programming, preferably in Fortran (all other programming environments are welcome), related to statistical algorithms, pseudo-random number generators, simulation techniques, and self-contained executable code to carry out new or interesting statistical methods.

Elegant derivations, as well as articles with no take-home message to practitioners, have low priority. Articles based on Monte Carlo (and other computer-intensive) methods designed to evaluate new or existing techniques or practices, particularly as they relate to novel applications of modern methods to everyday data analysis problems, have high priority.

Work appearing in *Regular Articles*, *Brief Reports*, and *Emerging Scholars* is externally peer reviewed, with input from the Editorial Board; work appearing in *Statistical Software Applications and Review* and *JMASM Algorithms and Code* is internally reviewed by the Editorial Board.

Please observe the following guidelines when preparing manuscripts:

1. *JMASM* uses a modified American Psychological Association style guideline.
2. Articles should be submitted without a title page or abstract. There should be no material identifying authorship except in the fields of the submission form. Include a statement in the cover letter indicating that proper human subjects protocols were followed where applicable, including informed consent.
3. Manuscripts should be prepared in Microsoft Word (.doc or .docx) only (Wordperfect and .rtf formats may be acceptable – please inquire). Please note that Tex (in its various versions), Exp, and Adobe .pdf formats are designed to produce the final presentation of text. They are not amenable to the editing process, and are NOT acceptable for manuscript submission.
4. The text maximum is 20 pages double spaced, not including tables, figures, graphs, and references. Use 11 point Times Roman font.
5. Create tables without boxes or vertical lines. Place tables, figures, and graphs "in-line", not at the end of the manuscript. Figures may be in .jpg, .tif, .png, and other formats readable by Adobe Illustrator or Photoshop.
6. The submission form requires an Abstract with a 50 word maximum, and a list of key words or phrases. Major headings are Introduction, Methodology, Results, Conclusion, and References. Center headings. Subheadings are left justified; capitalize only the first letter of each word. Sub-subheadings are left justified, indent optional.

7. Do not use underlining in the manuscript. Do not use bold, except for (a) matrices, or (b) emphasis within a table, figure, or graph. Do not number sections. Number all formulas, tables, figures, and graphs, but do not use italics, bold, or underline. Do not number references. Do not use footnotes or endnotes.
8. In the References section, do not put quotation marks around titles of articles or books. Capitalize only the first letter of books. Italicize journal or book titles, and volume numbers. Use "&" instead of "and" in multiple author listings.
9. Suggestions for style: Instead of "I drew a sample of 40" write "A sample of 40 was selected". Use "although" instead of "while," unless the meaning is "at the same time." Use "because" instead of "since," unless the meaning is "after." Instead of "Smith (1990) notes" write "Smith (1990) noted." Do not strike the spacebar twice after a period.

Journal of Modern Applied Statistical Methods

ISSN: 1538–9472

<http://digitalcommons.wayne.edu/jmasm>

PUBLISHED biannually (May, November) in partnership by:

JMASM, Inc.
PO Box 48023
Oak Park, MI 48237
ea@jmasm.com

Wayne State University Library System
Purdy Library
Detroit, MI 48202
digitalcommons@wayne.edu

Copyrights, Attribution and Usage Policies

Copyright ©2014 JMASM, Inc. *JMASM* retains the copyright for this work for the entire usual period, but grants assignors the right, after one year from the date of publication, to republish the work in whole or in part anywhere and in any format, provided reference is given to the original publication in *JMASM* (see website for further details). Readers may freely access journal content at <http://digitalcommons.wayne.edu/jmasm>.

To Advertisers

Advertisements are accepted at the discretion of the editor. Send requests for advertising information to ea@jmasm.com.

